Strategic Feedback Mechanisms and Experimentation*

Attila Ambrus[†], Haoyang Zhao[‡], Runxin Zhi[§]

August 18, 2026

Abstract

This paper studies how an information designer can influence experimentation with a new product of unknown quality. The designer commits to a history-contingent recommendation policy. A sequence of consumers arrives, not knowing their positions in the sequence or past outcomes, observing only the designer's recommendations. The designer learns from a noisy signal of quality generated from each purchase. When the designer maximizes expected consumer surplus, the first-best level of experimentation is achievable under general conditions. With alternative objectives—such as internalizing externalities on non-consumers or maximizing profit—the first best may fail, and at least one consumer incentive constraint must bind. We characterize the optimal second-best mechanisms under perfect good news and perfect bad news learning processes. With perfect good news, the designer can increase or decrease experimentation relative to the consumer-welfare benchmark until the relevant incentive constraint binds, and optimal policy may involve mixing before stopping. With perfect bad news, a designer with weaker experimentation incentives cannot do better than delaying experimentation, while a designer with stronger experimentation incentives cannot extend experimentation but may increase adoption by occasionally recommending a product already revealed to be of low quality.

*We are grateful to Johannes Hörner, Bruno Strulovici, Curtis Taylor, Fei Li, Rachel Kranton, Udayan Vaidya, participants at ERMAS 2026, the 37th Stony Brook International Conference on Game Theory, the Triangle Micro Conference 2026, and seminar participants at NYU Shanghai, Duke University and Duke Kunshan University, for their helpful comments.

[†]Department of Economics, Duke University. attila.ambrus@duke.edu

[‡]School of Economics, University of International Business and Economics. luke.zhao.du@outlook.com

[§]Department of Economics, Stony Brook University. runxin.zhi@stonybrook.edu

1 Introduction

Experimentation and adoption of new products and technologies have been investigated extensively in the Economics literature, both theoretically and empirically. A key issue in experimentation is that individual incentives typically do not align with the social value of experimentation. Experimentation with a new product can generate valuable information for future consumers, a positive externality that is typically not internalized by agents in the decision whether to experiment. This implies that starting experimentation with a product might not be individually rational, even when experimentation has a positive social value.

There is a line of literature, in both economics and computer science, on providing incentives to people to experiment.[1] Some of this literature examines incentives provided by information, both theoretically and empirically.[2] In particular, Kremer et al. (2014) and Che and Hörner (2018) show that an information designer (hereafter designer) who has some extra information relative to consumers about a new product can gradually kickstart experimentation through information design.[3] In this paper, we examine what such a designer can achieve when he does not have extra information beyond feedback from past consumers, in a context where the designed information structure determines how much a consumer knows about past purchases and consumer experiences. In particular, a consumer might not know (unless the information structure designed by the designer reveals it) how many other consumers purchased the product before her, that is, whether she is an early adopter or whether there is already a considerable amount of information generated about the product by past consumers.[4] Such environments naturally emerge when consumers receive private recommendations to buy products that might depend on past consumers' feedback about the product (on top of the given consumer's purchase history), and there are many new products arriving to the market, making it difficult for a consumer to judge exactly how recently a product was introduced. Examples of such private recommendation systems include *Amazon's Recommender*

[1]For recent contributions in computer science, see for example Immorlica et al. (2018), Mansour et al. (2022), Slivkins (2023) and Simchowitz and Slivkins (2024).

[2]For empirical investigations on how providing information can boost adoption of new technologies, see for example Hanna et al. (2014) and Ayalew et al. (2022). The latter paper considers targeted recommendations to different agents, which is a central feature of our analysis, too.

[3]More precisely, in Che and Hörner (2018) the designer learns about the product's quality from past consumers' experiences, the speed of which depends on the volume of consumption, but also from some independent sources that do not depend on consumption. See also Lyu (2026) for a related paper with a more general information structure in which the designer receives an initial private information about product quality, Lyu and Liao (2026) for a setting in which consumers can wait and stay around for multiple periods before deciding whether to purchase a product, and Madsen and Shmaya (2025) for how heterogeneity among agents more generally than differences in arrival time can limit what a designer can achieve.

[4]Gershkov and Szentes (2009) make a similar assumption in a different context, with costly information acquisition and voting by a sequence of agents.

System, and recommendations sent by email by *AliExpress*, featuring relatively new products like electronic gadgets and games. A different scenario is a physician deciding whether to recommend a treatment that has not yet received FDA approval to patients, who do not know how much feedback the physician already observed about the effects of the medical substance.[5]

Our model features a sequence of consumers (she) arriving in the marketplace, each deciding whether to purchase a good of unknown quality. If the product is good, the consumer gets a utility of 1 when purchasing it, and if it is bad, she gets a utility of -1. Not purchasing the product gives a sure utility of 0. The consumers have a common prior belief regarding the quality. After a consumer arrives, there is a $1-\delta$ probability that consumers potentially interested in the product stop arriving, for example, because the product becomes obsolete, so the total number of arriving consumers is random. An arriving consumer does not know her place in the sequence; instead, she can only draw inferences about it.

If a consumer purchases the product, she receives a signal about its quality and then leaves a feedback about her experience.[6] For the recommendations received by different consumers, we employ an information design approach. In particular, the designer can publicly commit to a recommendation mechanism that recommends the purchase of the product with some probability as a function of the history of previous purchases and feedback.

We first demonstrate how the designer can use a recommendation mechanism to affect experimentation in a two-period example. Suppose that the prior belief that the product is good is less than $1/2$, but not by much. Suppose also that if the product is good, a consumer who purchases it has a good experience (thereby proving that the product is good) with high probability. Without a recommendation mechanism, consumers would not purchase the product, since starting experimentation is not individually rational. One way the designer can kickstart experimentation is by recommending purchase to the consumer arriving in Period 1 and to the consumer arriving in Period 2 only if the first consumer left positive feedback. A consumer who is recommended to buy the product knows that with some probability she is the very first consumer (in which case she would rather not buy the product), but she also assigns some probability to the event that someone has already tested the product and verified it to be good, which makes her more willing to follow the recommendation to purchase the good.

[5]For instance, a physician may prescribe an FDA-approved drug for an off-label use, namely, a condition or population beyond the scope of its FDA approval. Moreover, many medical procedures and surgical techniques are not subject to FDA approval in the same manner as drugs, and their adoption often relies on clinical evidence, professional standards, and physician judgment.

[6]For simplicity, we assume that leaving feedback is costless and that consumers leave truthful feedback about their experience with the product. In reality, leaving feedback might be driven by other factors, including preferences toward future consumers, a desire for bad products to give way to good products if there is a possibility of repeat purchasing needs in the future, or platforms providing monetary incentives, such as randomly awarded gift certificates, to consumers leaving feedback.

The central focus of our investigation is how close the designer can get to achieving his first-best experimentation policy, via recommendation mechanisms. This of course depends on what the objective function of the designer is. Our first result is that if the designer's goal is to maximize the expected intertemporal consumer surplus, that is, the sum of expected payoffs of consumers discounted by arrival probabilities, then the designer can achieve the first best for general specifications of the model. This simply follows from the fact that for any recommendation mechanism, the expected payoff of a consumer from purchasing conditional on a purchase recommendation is a fixed fraction of the designer's payoff resulting from consumers following the recommendations of the mechanism. That is, the designer's and the consumers' payoffs go hand in hand. In particular, if the optimal experimentation plan generates a strictly positive consumer surplus, and if the recommendation mechanism follows the plan that maximizes consumer surplus, a consumer who is recommended to purchase the good expects a strictly positive payoff when following the recommendation. Intuitively, not knowing her position in the sequence of arrivals puts a consumer behind the veil of ignorance, helping to align individual and social incentives.

In general, a recommendation mechanism needs to satisfy two Incentive Compatibility (IC) constraints. One is that if a consumer is recommended to purchase the product, her payoff from the purchase has to be nonnegative. The second one is that if she is not recommended to purchase the product then her payoff from buying it is nonpositive. Since consumers do not know their position in the sequence of arrivals, these constraints are not the standard local constraints, pertaining to a consumer arriving in a certain period, but global, meaning that they depend on whether purchasing on average given a certain recommendation yields a positive payoff or not. This means that the techniques needed to solve the designer's problem are different from those in the related literature.

For general objective functions of the designer, and for general learning signal structures from experimentation, we show that either the designer can achieve his first best, or exactly one of the two IC constraints needs to bind at the second-best optimal recommendation mechanism of the designer. We also show that without loss of generality, the second-best recommendation policy has a threshold form, such that above a certain posterior belief the designer recommends buying the product, and below this threshold he does not. When the first best is not IC, mixing may be necessary to keep the corresponding IC constraint binding and only happens at the threshold belief.

The existence of a threshold second-best recommendation policy, while intuitive in the stationary environment we work with, is far from obvious as changing the probability of recommending purchase at a given history changes the probabilities of reaching subsequent histories associated with different posterior beliefs, which can affect both global IC constraints in complicated ways. Hence, starting from a non-threshold policy, it might not be possible to find a local modification of the policy that respects the IC constraints and improves the designer's expected payoff. We overstep the problem by rewriting the designer's problem in a form in which the IC constraints are built into the objective function and show that in this unconstrained auxiliary problem there is an optimal policy that has the threshold form.

While there always exist threshold second-best policies, the specific second-best policies under different signal structures can look rather different. For a more concrete characterization of the second-best recommendation mechanism, we investigate two particular learning signal structures popular in the literature, perfect good news and perfect bad news. Perfect good news means that after purchasing the product, a consumer can either get a good signal or no signal, and there is a positive probability of getting a good signal only if the product is good. An example is a medicine that has a positive probability of curing the patient if it is good, but zero possibility of curing the patient if it is bad. Perfect bad news, on the other hand, means that a consumer buying the product can either receive a bad signal or no signal, and the bad signal only has a positive probability when the product is bad. An example would be a television that has some probability of exploding if it is a bad product, but not if it is a good product. For both of these learning signal structures, we investigate cases when the designer prefers experimentation beyond the level that maximizes consumer surplus (over-experimenting designer) and cases when the designer prefers less experimentation (under-experimenting designer). One reason why the designer's goals might be different from maximizing consumer surplus is the existence of positive or negative externalities from agents consuming the product. An extreme case of an over-experimenting designer is that of a profit maximizing seller, who only cares about the total amount of sales, independent of whether the product is good or bad. On the other extreme, a designer might really want to discourage consumers to buy a product like an addictive substance, so in the first best he would want to recommend no purchase to every consumer. But if given only this generic do-not-purchase recommendation the product is too tempting to acquire for consumers, the designer needs to recommend purchase in some instances, in order to make a do-not-purchase recommendation IC.

In the perfect good news case, if the over- or under-experimentation incentives of the designer are not too strong, he can still implement his first best under most prior beliefs. When these incentives are stronger, however, the first best violates one of the IC constraints, and the second-best policy adjusts experimentation only until the relevant IC constraint binds. The optimal policy consists of an exploration phase followed by an exploitation phase. During the exploration phase, the designer recommends purchase as long as no good news has arrived, possibly mixing in the final period of experimentation to make the binding IC constraint hold exactly. If good news arrives, the product is revealed to be good and the designer enters the exploitation phase, recommending purchasing a known good product thereafter. If no good news arrives by the end of the exploration phase, the designer stops recommending purchase. In addition, the second-best policy does not change with respect to the experimentation incentives once regulated by the binding IC constraints.

The perfect bad news case is qualitatively different. Here we show that an over-experimenting designer cannot do much as long as he would not want purchases of a bad product. In this case if the consumer surplus is maximized by no experimentation then the designer's second best involves no experimentation, and if the consumer surplus is maximized by recommending the product until bad news arrives then the over-experimenting designer cannot do better than implementing that. Things change when the designer would like purchases even in the bad state, such as in the case of the designer being a profit maximizing seller. In this case the seller might recommend purchase even in some cases when bad news arrives and the low quality is revealed to the seller. And for an under-experimenting designer, we show that if he cannot achieve no experimentation then his potential impact on experimentation is very limited: he cannot do better than postponing the start of experimentation by one period with some probability.

The different forms of the second-best policies under perfect good news and perfect bad news, despite their common threshold structure, arise from the opposite directions in which beliefs evolve when no news arrive. Under perfect good news, the absence of good news makes the product progressively less likely to be good. Experimentation therefore becomes less attractive over time, generating an exploration phase in which purchase is recommended until the belief reaches a stopping threshold, followed by an exploitation phase only if good news has revealed the product to be good. The designer then in general can adjust the duration of experimentation, in either direction. Under perfect bad news, by contrast, the absence of bad news makes the product progressively more likely to be good, so experimentation becomes easier to sustain over time. If the designer does not value sales of a product revealed to be bad, no purchase is recommended after bad news arrives. In this case, an over-experimenting designer cannot improve on the consumer-efficient policy: before bad news, consumers already experiment whenever doing so can be made incentive compatible, while after bad news neither the designer nor consumers wish to continue. For an under-experimenting designer, if consumers are willing to experiment at the prior, a policy that always recommends no purchase would violate the incentive constraint associated with the no purchase recommendation. Incentive compatibility then implies that purchase needs to be recommended with positive probability at the prior belief, and with probability 1 at all higher beliefs.

We derive most of our results through reformulating the problem of the designer in terms of occupancy measures, that is, the expected number of periods that the game is in a certain state, namely that at the beginning of the period the belief about the product being good is at a certain value, and the expected number of periods that at that belief purchase is being recommended. This representation summarizes the long-run behavior of a recommendation policy in terms of state-action frequencies, which simplifies both the designer's objective function and the consumers' IC constraints. In particular, the designer's problem can be reformulated as a linear programming problem in occupancy measures. Moreover, typically the designer's second-best policy is unique with respect to occupancy measures, but not in terms of recommendation policy, making the characterization in terms of occupancy measures cleaner.

In an extension, we also examine the relaxation of the assumption that a consumer does not know her place in the sequence of arrivals in a simple version of the model. In particular, we assume that a consumer, besides receiving a recommendation from the mechanism, receives a noisy signal about her place in the arrival sequence. We show that if this signal is not too informative then our conclusions remain valid. In a different extension, we investigate the impact of not all consumers leaving feedback. We show that this leads to a slower learning technology about product quality, reducing the incentives of the designer to facilitate experimentation, but otherwise leaves our qualitative conclusions intact.

Both Kremer et al. (2014) and Che and Hörner (2018) consider extensions with consumers having less information about their relative arrival time, in the context when the designer aims to maximize expected consumer surplus. In Kremer et al. (2014) there are two goods, initially both with unknown quality, and every consumer buys one of the goods. A simple learning technology is adopted, in which consuming a good just one time immediately reveals its quality. For the particular learning technology of the model the paper argues that if consumers know little information about their positions in the sequence then the first best can be achieved, which we formally show for general learning processes. Che and Hörner (2018) formally show this result for the case of perfect good news and perfect bad news (the latter in an appendix) for high enough levels of patience, so our result for the particular case of a designer wanting to maximize expected consumer surplus is a generalization of their finding for all (stationary) learning processes and levels of patience.

Kovbasyuk and Spagnolo (2024) also consider recommendation systems in a Markovian environment. Although their setting is different in that the quality of the product can change over time, and in that the current quality is perfectly revealed with each purchase, the optimal recommendation policy is of the threshold form, like in our setting. This result is more straightforward to obtain in their setting, in which the planner's belief is fully determined by the last feedback left and the time that has lapsed, compared with the general (stationary) information processes that we work with.

More generally, our paper builds on the literature on strategic experimentation (Bolton and Harris (1999), Keller et al. (2005), Keller and Rady (2010, 2015)), information design (see Rayo and Segal (2010) and Kamenica and Gentzkow (2011), and for a more recent survey Bergemann and Morris (2019)), and information cascades (see Banerjee (1992), Bikhchandani et al. (1992), and for a more recent survey Bikhchandani et al. (2024)). Papers in the latter literature assume that the information of consumers about previous purchases and consumer experiences are exogenously given, while in this paper we assume that they are endogenously determined by the designer's choice of the recommendation mechanism.

2 The Model

Time is discrete and infinite, $t = 1, 2, \cdots$. There is a product with persistent quality q that is either good or bad. That is, $q \in \Omega \equiv \{G, B\}$. The potential consumers of the product arrive sequentially, and let consumer t denote the consumer who arrives at time t. Consumer 1 arrives for sure. If consumer t arrives, consumer $t+1$ arrives with probability $\delta \in (0, 1)$. If consumer $t+1$ does not arrive then no more consumers arrive afterward. A consumer does not know the time of her arrival, and her belief regarding her arrival time follows a geometric distribution with parameter $1-\delta$. That is, she believes that she is consumer t with probability $\delta^{t-1}(1-\delta)$. All consumers share the common prior μ_0 that the product is good. A consumer's payoff is 1 if she purchases a good product, -1 if she purchases a bad product, and 0 if she does not purchase. If consumer t purchases the product, by using it she receives a signal about the quality

$$\pi : \Omega \to \Delta(S),$$

where S is the set of all possible signal realizations. We assume that the signal is independent across consumers, conditional on the state.

In this context we consider the problem of a designer who gets to observe the signal realizations of consumers who buy the product, via feedback that these consumers leave after using the product, and can commit to a recommendation mechanism that can recommend a newly arriving consumer to either purchase (Y) or not to purchase (N) the product, as a function of the history of signals of previous consumers.[7] At the beginning of the game, the designer also shares the same prior μ_0. For each consumer, the designer gets

$$v_t = \begin{cases} 1 & \text{if she purchases a good product,} \\ 0 & \text{if she does not purchase,} \\ a & \text{if she purchases a bad product,} \end{cases}$$

where $a \in (-\infty, 1]$ is a parameter. The designer's total expected payoff is

$$\sum_{t=1}^{\infty} \delta^{t-1} \mathbb{E}(v_t).$$

A special case is $a = -1$, where the designer's payoff regarding a consumer coincides with the consumer's payoff. This is referred to as the *nonpaternalistic* baseline. If $a > -1$, the designer has a stronger experimentation incentive. In particular, if $a > 0$, the designer always prefers a

[7] In general the designer could generate a richer set of messages for the consumers, but because the action set of a consumer is binary, standard revelation principle arguments establish that it is without loss of generality to focus on structures with two messages, with meanings of recommending or not recommending purchase.

consumer to purchase the product, even if the product is of bad quality. If $a = 1$, the designer can be considered as a profit-maximizing seller trying to maximize sales of the product, independent of the state. On the other hand, if $a < -1$, the designer has a weaker experimentation incentive. In the limit case of $a \to -\infty$, the designer never wants any consumer to purchase, unless the product is known to be of good quality.

A history includes all previous recommendations, all previous consumer purchasing decisions, and all realized signals. In general, the designer's strategy maps any history to a recommendation.

2.1 A Motivational Example

We start with a simple example to show how the designer can facilitate experimentation. For this example suppose that there are only two periods. Consumer 1 arrives for sure, while Consumer 2 arrives with probability δ. The signal is *perfect good news*,

$$\begin{bmatrix} \mathbb{P}(g \mid G) & \mathbb{P}(n \mid G) \\ \mathbb{P}(g \mid B) & \mathbb{P}(n \mid B) \end{bmatrix} = \begin{bmatrix} \alpha & 1-\alpha \\ 0 & 1 \end{bmatrix}.$$

In other words, if the state is G, the good news g arrives with probability α. Otherwise, there is no news (denoted by n). Let $a = -1$ in this example. Suppose $\mu_0 < 1/2$ so that persuading consumers to purchase is not trivial.

We solve the designer's first best by backward induction. If Period 2 is reached, the designer should recommend Y if and only if g is received from Consumer 1's purchase. To see this, notice that if Consumer 1 did not purchase, or b is received, the designer holds a belief at most μ_0. Since there is no further consumer (the game ends there), recommending a product that is more likely to be bad is not beneficial. If g is received, the product is known to be good, so that Y should be recommended.

Given this, recommending Y in the first period yields

$$\underbrace{\mu_0 \times 1 + (1-\mu_0) \times (-1)}_{\text{Immediate Payoff}} + \underbrace{\delta\alpha\mu_0 \times 1}_{\text{Experimentation Benefits}}.$$

The designer should recommend Y in the first period if this is nonnegative, since no purchase yields 0. This gives

$$\mu_0 \geqslant \frac{1}{2+\delta\alpha} = \underline{\mu}_0. \tag{1}$$

Notice that $\underline{\mu}_0 < 1/2$. Therefore, the designer's first-best recommendation policy is as follows:

- If $\mu_0 \geqslant \underline{\mu}_0$, recommend Y at Period 1. If g is received, recommend Y at Period 2. Otherwise, recommend N at Period 2.
- If $\mu_0 < \underline{\mu}_0$, recommend N at both periods.

It turns out that this recommendation mechanism is also incentive compatible. It is easy to see that the consumers follow the recommendation N, which only happens when the designer's belief is strictly below 1/2. When the designer recommends Y, consumers cannot tell whether they are at the first or the second period. The expected belief is then

$$\frac{1}{1+\delta\mu_0\alpha} \times \mu_0 + \frac{\mu_0\alpha\delta}{1+\delta\mu_0\alpha} \times 1 = \frac{1+\mu_0\alpha\delta}{1+\delta}.$$

A consumer chooses to follow the recommendation Y if this is at least 1/2, which is equivalent to $\mu \geqslant \underline{\mu}_0$. We will show that this is not a coincidence: When $a = -1$, the recommendation mechanism corresponding to the designer's first-best experimentation plan is incentive compatible.

3 The designer's First Best

This section considers the designer's first-best problem, in which the consumers always follow the designer's recommended action. Clearly, if $a \geqslant 0$, the designer should always recommend Y, since purchasing is always better than not purchasing. In what remains, we consider the case that $a < 0$.

Suppose that μ_t is the belief of the designer at the beginning of period t and consumer t arrives. Let the designer recommends Y, and let $s \in S$ be the realized signal. The posterior belief is

$$\mu_{t+1} = \frac{\mu_t \pi(s \mid G)}{\mu_t \pi(s_t \mid G) + (1-\mu_t)\pi(s \mid B)}. \tag{2}$$

(2) shows the distribution of the next state depends on history only through μ and the current action. Also, the signal structure π is stationary, the action set $\{Y, N\}$ is fixed at each state μ, the designer's payoff depends on μ and current action only, and the consumer arrival process is memoryless. By standard results (Stokey and Lucas, 1989; Puterman, 1994), the designer's problem is a stationary Markov decision problem with state μ, and there exists an optimal policy that is Markov, $r : [0,1] \to \Delta\{Y, N\}$.

The designer's Bellman equation is

$$W(\mu) = \max\{\delta W(\mu), (1-a)\mu + a + \delta \mathbb{E}_\pi[W(\mu') \mid \mu]\}.$$

If $\delta W(\mu) > (1-a)\mu + a + \delta\mathbb{E}_\pi[W(\mu')]$, then $W(\mu) = \delta W(\mu)$, so that $W(\mu) = 0$. Intuitively, if the designer finds it optimal not to experiment now, the belief remains unchanged, and the designer (under Markov strategies) stays away from experimentation in all future periods. The designer chooses to experiment for sure if the continuation payoff $(1-a)\mu + a + \delta\mathbb{E}_\pi[W(\mu')]$ is strictly positive. Here, $(1-a)\mu + a$ is the direct payoff from purchasing the product at the current period, and $\delta\mathbb{E}_\pi[W(\mu')]$ is the option value from experimenting. In general, the designer may still be willing to experiment if the direct payoff is negative, since experimenting now helps the designer learn about the quality of the product and make better recommendations in all future periods. The designer can choose a mixed strategy if the continuation payoff is exactly 0.

We next show that there always exists a designer's first-best policy that is a threshold policy.

Theorem 1 (Threshold Policy)**.** There exists a threshold $\underline{\mu} \in (0,1)$ such that the designer recommends Y for sure if $\mu > \underline{\mu}$ and N for sure if $\mu < \underline{\mu}$.

Proof. We first show that π satisfies first order stochastic dominance. That is, for $\mu_1 < \mu_2$, the posterior distribution of μ under π with prior μ_2 first order stochastically dominates the posterior distribution of μ under π with prior μ_1. Let G_1 and G_2 be the CDF of the posterior distribution when the priors are μ_1 and μ_2, respectively. We want to show that for any $x \in (0,1)$, $G_1(x) \geqslant G_2(x)$. Pick $x \in (0,1)$. If x is a posterior following prior μ_2, then for some signal s,

$$x = \frac{\mu_2 \pi(s \mid G)}{\mu_2 \pi(s \mid G) + (1-\mu_2)\pi(s \mid B)} = \left[1 + \frac{1-\mu_2}{\mu_2}\frac{\pi(s \mid B)}{\pi(s \mid G)}\right]^{-1}.$$

Therefore, if we define

$$S_x^2 \equiv \left\{ s \in S : \frac{\pi(s \mid B)}{\pi(s \mid G)} \geqslant \frac{1-x}{x}\frac{\mu_2}{1-\mu_2} \right\},$$

which is the set of signals that leads to a posterior at most x under the prior μ_2, then

$$G_2(x) = \mu_2 \mathbb{P}_\pi(s \in S_x^2 | G) + (1-\mu_2)\mathbb{P}_\pi(s \in S_x^2 | B),$$

where $\mathbb{P}_\pi(s \in S_x^2 | G)$ is the probability that $s \in S_x^2$ given the signal structure π and quality state G, similarly for $\mathbb{P}_\pi(s \in S_x | B)$.

Since $\mu_1 < \mu_2$, if we define S_x^1 as the set of signals that leads to a posterior at most x under the prior μ_1, $S_x^2 \subseteq S_x^1$, since for each given signal, the posterior of that signal under a higher prior is higher. It follows that

$$\begin{aligned} G_1(x) &= \mu_1 \mathbb{P}_\pi(s \in S_x^1 | G) + (1-\mu_1)\mathbb{P}_\pi(s \in S_x^1 | B) \\ &\geqslant \mu_1 \mathbb{P}_\pi(s \in S_x^2 | G) + (1-\mu_1)\mathbb{P}_\pi(s \in S_x^2 | B). \end{aligned}$$

Therefore,

$$G_1(x) - G_2(x) \geqslant (\mu_2 - \mu_1)\left[\mathbb{P}_\pi(s \in S_x^2 | B) - \mathbb{P}_\pi(s \in S_x^2 | G)\right],$$

and the sign is determined by the terms in the bracket. Notice that S_x^2 gathers all the signals with an (inverse) likelihood ratio $\pi(s|B)/\pi(s|G)$ greater than the threshold $(1-x)\mu_2/x(1-\mu_2)$. Since $\mathbb{P}_\pi(s \in S|B) - \mathbb{P}_\pi(s \in S|G) = 1 - 1 = 0$, gathering the signal with larger inverse likelihood ratio must make this difference larger. This shows the terms in the bracket is positive, so that FOSD indeed holds.

We now show the main claim by showing that W is increasing in μ. Let T be the Bellman operator, i.e.,

$$(Tf)(\mu) = \max\{0, (1-a)\mu + a + \delta \mathbb{E}_\pi[f(\mu') \mid \mu]\},$$

for $f : [0,1] \to \mathbb{R}$. We first show that T preserves monotonicity. That is, if f is an increasing function, so is Tf.

Again let $\mu_1 < \mu_2$ be given. We want to show $Tf(\mu_1) \leqslant Tf(\mu_2)$. There are two cases:

- $Tf(\mu_1) = 0$. Then, since $Tf(\mu_2) \geqslant 0$, $Tf(\mu_1) \leqslant Tf(\mu_2)$ holds.
- $Tf(\mu_1) > 0$. Consider $(1-a)\mu + a + \delta \mathbb{E}_\pi[f(\mu') \mid \mu]$. Since $a < 0$, $(1-a)\mu_1 + a < (1-a)\mu_2 + a$. Since π satisfies FOSD and $f(\cdot)$ is an increasing function, $\delta \mathbb{E}_\pi[f(\mu') \mid \mu_1] \leqslant \delta \mathbb{E}_\pi[f(\mu') \mid \mu_2]$.

Therefore, T preserves (increasing) monotonicity as desired.

Define $W_0(\mu) = 0$ for all μ. Let $W_{n+1}(\mu) = TW_n(\mu)$ for all μ and $n \geqslant 1$. Each W_n is an increasing function since T preserves monotonicity. By Blackwell's contraction mapping theorem, T is a contraction mapping, and $W_n \to W$, which shows W is an increasing function.

Suppose that at some μ_3 the designer recommends Y with positive probability. It follows that

$$W(\mu_3) = (1-a)\mu_3 + a + \delta \mathbb{E}_\pi[W(\mu') \mid \mu_3] \geqslant 0.$$

Consider any $\mu_4 > \mu_3$. Similar to the argument above, we know that $(1-a)\mu_4 + a > (1-a)\mu_3 + a$ (since $a < 0$) and $\delta \mathbb{E}_\pi[W(\mu') \mid \mu_4] \geqslant \delta \mathbb{E}_\pi[W(\mu') \mid \mu_3]$ (by FOSD and monotonically increasing W). This shows

$$(1-a)\mu_4 + a + \delta \mathbb{E}_\pi[W(\mu') \mid \mu_4] > 0,$$

which in turn shows that $W(\mu_4) = (1-a)\mu_4 + a + \delta \mathbb{E}_\pi[W(\mu') \mid \mu_4] > 0$ and Y is recommended for sure. Therefore, if Y is recommended with positive probability at some belief, then Y should be recommended for sure at all higher beliefs. □

4 The Designer's Second Best

We now consider the designer's second-best problem, where the consumers' incentive compatibility constraints must be respected. As discussed previously, by the revelation principle, the designer faces two incentive compatibility constraints: the incentive compatibility constraint after a recommendation Y (hereafter ICC-Y), and the incentive compatibility constraint after a recommendation N (hereafter ICC-N). For ICC-Y, it means the consumers' expected payoff of purchasing after observing a recommendation Y must be weakly positive. Equivalently, the consumers' posterior belief after observing a recommendation Y is at least 1/2. ICC-N can be expressed similarly.

In this section, we start from the nonpaternalistic baseline, i.e., a designer with $a = -1$. We will show that the nonpaternalistic designer can always implement his first best even with the IC constraints. We then consider paternalistic designers who have over- or under-experimentation incentives (i.e., $a > -1$ or $a < -1$). In those cases, we show that the designer's second-best policies still maintain a threshold form, and mixing at the threshold is necessary in general for designers with strong over- or under-experimentation incentives.

4.1 Nonpaternalistic Baseline

We first consider the case when $a = -1$. This means the stage payoff of the designer under belief μ is $\mu \times 1 + (1-\mu) \times (-1) = 2\mu - 1$, which coincides with the consumers' payoff. In this section, we will show that the designer with $a = -1$ can always achieve the first best even if consumers can choose whether or not to follow the designer's recommendation.

To establish this result, we introduce the concept of **occupancy measures** over beliefs, which significantly simplifies the analysis and underpins many of our subsequent findings. Rather than tracking entire histories, occupancy measures capture the expected number of periods in which each belief is held along the path, as well as the expected frequency with which action Y is recommended under each belief. This formulation enables us to recast the designer's problem as a linear programming problem. Formally, define the following notations:

- Let $x(\mu)$ be the expected number of periods the belief at the beginning of the period is μ, and there is a consumer arriving.
- Let $y(\mu)$ be the expected number of periods for which the belief at the beginning of the period is μ, there is a consumer arriving, and Y is recommended.

The occupancy measures depend on the prior, signal structure, and the designer's policy.

Theorem 2. The designer's first-best recommendation policy is incentive compatible.

Proof. We will show that ICC-Y and ICC-N are both satisfied in the first-best policy if $a = -1$. We start with ICC-N. Notice that when N is recommended in the first best, the designer must have a belief at most $1/2$. In other words, $r(\mu) > 0$ if $\mu > 1/2$. To see this, suppose the contrary that at some $\mu > 1/2$, $r(\mu) = 0$ in the first best. But then the designer has a profitable deviation by recommending Y for sure at μ at this period and recommend N at all future periods regardless of the realized signal. This guarantees a strictly positive expected payoff, while $r(\mu) = 0$ yields a zero payoff for sure. This shows $r(\mu) = 0$ cannot be a first best. Therefore, whenever N is recommended, consumers know that the designer's belief is at most $1/2$, and following the recommendation N by not purchasing is incentive compatible.

Next, consider ICC-Y. Assume that Y is recommended with positive probability, which means $W(\mu_0) \geqslant 0$. The goal is to show that the consumer's expected posterior payoff after observing Y is proportional to $W(\mu_0)$.

Given the first-best policy is Markov, the occupancy measure is always well-defined and depends on the prior. Specifically,

$$
\begin{aligned}
y(\mu) &= r(\mu)x(\mu), \\
x(\mu) &= \mathbb{I}(\mu = \mu_0) + \delta \int y(\hat{\mu}) \mathrm{d}\pi(\mu \mid \hat{\mu}) + \delta[1 - r(\mu)]x(\mu).
\end{aligned}
$$

For $y(\mu)$, $r(\mu)$ is the first-best policy of recommending Y at belief μ. For $x(\mu)$, the first term considers whether or not μ is the prior, given that the prior is where the game starts and naturally have one more period in the occupancy. The second term considers the cases where an experiment at previous period yields a signal that leads to posterior μ. The last term considers the cases that no experiment was recommended and the belief remains constant for the next consumer. The designer's expected payoff under the first-best policy is

$$
W(\mu_0) = \int y(\mu)(2\mu - 1) \mathrm{d}\mu \geqslant 0
$$

since $r(\mu_0) > 0$.

On the other hand, the consumer's posterior belief after observing Y is described by

$$
\int \frac{y(\mu)\mu}{\int y(\hat{\mu}) \mathrm{d}\hat{\mu}} \mathrm{d}\mu,
$$

and the consumer purchases if this is at least $1/2$. This means

$$
2 \int y(\mu)\mu \geqslant \int y(\hat{\mu}) \mathrm{d}\hat{\mu}.
$$

Move the RHS to the LHS, use the same integral variable, and collect terms to get

$$
\int y(\mu)(2\mu - 1) \mathrm{d}\mu \geqslant 0,
$$

which was our starting assumption. □

Instead of working directly with the probabilities of recommending Y after different histories, in the proof we employ a strategy of working with occupancy measures of states in the game with different starting beliefs at the beginning of the period and purchases of the good with such beliefs.[8] The idea is as follows. Ultimately, the designer and the consumers only care about how often Y is recommended at different beliefs, once these frequencies are adjusted according to the arrival probabilities of the consumers. For example, a recommendation plan is IC if Y is recommended often when the product is good with high probability and less often when it is more likely to be bad.

[8] For a classic reference on occupancy measures, see Puterman (1994).

It is intuitive why the first-best policy is IC when $a = -1$. Since the designer does not hold any prior information advantage, when knowing the designer's recommendation policy, each consumer can fully trace out the potential belief evolution the same way as the designer at the beginning of Period 1. When the designer recommends N, the designer gets 0. When he recommends purchasing, he gets the same expected stage payoff as the corresponding consumer. Therefore, a consumer's expected payoff is just a scaled version of the value of the designer under the prior belief.

Example 1. As an example, we consider the perfect-good-news signal again. This simple form of the signal structure allows us to solve for the closed form of $\underline{\mu}$. At $\underline{\mu}$, if the designer recommends purchasing, either g is realized with probability $\underline{\mu}\alpha$, and the belief jumps to 1, or n is realized with probability $1 - \underline{\mu}\alpha$, and the belief falls below $\underline{\mu}$. At the boundary, using the Bellman equation,

$$0 = W(\underline{\mu}) = 2\underline{\mu} - 1 + \delta \left[\underline{\mu}\alpha \frac{1}{1-\delta} + (1 - \underline{\mu}\alpha) \cdot 0 \right],$$

which gives

$$\underline{\mu} = \frac{1-\delta}{2(1-\delta) + \delta\alpha}. \tag{3}$$

If $\mu_0 > \underline{\mu}$, the designer recommends for a fixed number of periods. If there is good news arriving in these periods, the quality is revealed to be G, and the designer recommends purchasing the product forever. If no good news arrives for enough number of periods, the belief drops below $\underline{\mu}$, and the designer recommends N in all future periods. ■

4.2 Paternalistic Designer

We now consider designers with over- ($a > -1$) or under- ($a < -1$) experimentation incentives. The designer's problem differs from a standard stopping time problem due to the IC constraints. Specifically, the IC constraint is a global constraint[9], not local, as consumers are time-blind.

With the IC constraint, the first-best policy may or may not be implementable for such designers. For example, for $a > -1$, the designer in the first best sometimes would like consumers to experiment even when it decreases expected consumer surplus. This decreases the expected payoff of a consumer when getting a Y recommendation. If the resulting expected payoff is still above 0, the first best is implementable. But for high enough a, the first-best policy of the designer overrecommends purchase to the extent that the expected payoff conditional on a Y recommendation becomes negative, meaning that the first best becomes unattainable. The following result formalizes this argument.

Proposition 3. *Suppose $a > -1$ ($a < -1$) and the designer's first-best policies cannot be implemented. Then in any second-best policy, ICC-Y (N) is binding, while ICC-N(Y) is slack.*

[9]The constraints are global in the sense that the complete recommendation policy is involved, and the payoffs are aggregated over time in these constraints.

Proof. For simplicity of notation, we let $\gamma^r(Y)$ and $\gamma^r(N)$ be the posterior beliefs of consumers after observing recommendation Y and N, respectively, under recommendation policy r. In what follows, we will use r' and r'' to denote a first-best and a second-best recommendation policy in a given game.

Case 1: $a > -1$. Suppose that the designer's first-best policies are not implementable. If the designer commits to a second-best policy in which Y is never recommended, any claim regarding ICC-Y, including that it is binding, is vacuously true. Here, ICC-N is indeed slack, since never recommending Y cannot be the second best if $\mu_0 \geqslant 1/2$.[10] Given $\mu_0 < 1/2$ and N is always recommended, $\gamma^{r''}(N) = \mu_0 < 1/2$.

In what follows, we consider the cases when Y is recommended with positive probability after some history in a second-best policy r''.

First, we argue that ICC-N cannot bind in r''. Otherwise, $\gamma^{r''}(N) = 1/2$. Since a second-best policy must be IC, $\gamma^{r''}(Y) \geqslant 1/2$. Bayesian consistency then implies $\mu_0 \geqslant 1/2$. But this suggests a first-best policy is implementable.

Next, we argue that ICC-Y must bind in r''. Suppose the contrary that $\gamma^{r''}(Y) > 1/2$. Define a new recommendation policy r_1 as follows:

- At the beginning of Period 1, the designer randomizes between r'' and a first-best policy r'. The designer plays r' with probability $\varepsilon > 0$.

The designer gets higher expected payoff under r_1 (since the first-best payoff is strictly higher). For sufficiently small ε,

$$\gamma^{r_1}(Y) = (1-\varepsilon)\gamma^{r''}(Y) + \varepsilon\gamma^{r'}(Y) \geqslant 1/2.$$

And we know that $\gamma^{r_1}(N) < 1/2$ since both $\gamma^{r'}(N)$ and $\gamma^{r''}(N)$ are less than $1/2$. Therefore, r_1 is also IC. This shows that the "second-best policy" is in fact not the second best, a contradiction.

Case 2: $a < -1$. If N is never recommended in a second-best policy, ICC-N binding is again vacuously true. To see that ICC-Y is slack, note that the designer recommends Y at the first period if the ex ante expected payoff at μ_0 is weakly positive. Since $a < -1$, an (imaginary) designer with $a = -1$ under the same recommendation policy must have a strictly positive ex ante expected payoff at μ_0. By Theorem 2, the consumers payoff after observing recommendation Y has the same sign as the ex ante payoff of the designer with $a = -1$, so that the sign is strictly positive. This shows ICC-Y is slack.

Suppose that N is recommended with positive probability after some history in a second-best policy r''. We first argue that ICC-Y is slack. If Y is never recommended, it is again vacuously true that ICC-Y is slack. If Y is recommended with positive probability, it must be recommended with positive probability at Period 1. Then the argument above again shows ICC-Y is slack.

[10] In fact, when $a > -1$ and $\mu_0 \geqslant 1/2$, the designer can always implement a first-best policy: since $a > -1$, $\gamma^{r'}(N) < 1/2$ for any first-best policy r'. Bayesian consistency suggests that $\gamma^{r'}(Y) > 1/2$ since $\mu_0 \geqslant 1/2$.

Finally, we show that ICC-N must bind. Suppose not, that is, both ICC-Y and ICC-N are slack in the second-best policy. Similarly to the cases of $a > -1$, the designer can adopt a new recommendation policy by randomizing between the second-best policy (with a large probability) and a first-best policy (with a small enough probability). The new policy yields a strictly higher payoff and is IC. Therefore, the "second-best policy" is in fact not a second-best policy, which is a contradiction. □

Intuitively, when $a > -1$ and the first best is not IC, the designer has sufficiently strong incentives to over-experiment. ICC-Y now acts as the designer's budget constraint: leaving ICC-Y slack means the designer could recommend Y at more histories, decrease the posterior after recommendation Y, and improve expected payoffs. Similarly, when $a < -1$ and the first best is not IC, ICC-N is the binding budget constraint. Leaving ICC-N slack means that the designer could recommend N in more histories, increase the posterior after recommendation N, and improve expected payoffs.

With the global IC constraints, the designer's problem differs from a standard optimal stopping problem. Nevertheless, below we show that there always exists a second-best policy that has the threshold belief feature, and mixing, if necessary, only happens at the threshold belief. The proof of this result is much more involved than that of Theorem 1 (the existence of a first-best policy that has the threshold form) because the global IC constraints can connect histories that are predecessors and successors of each other, and so changing the recommendation probability at a certain history can affect both IC constraints in complicated ways, via changing the probabilities of subsequent histories being reached.

Theorem 4. For any $a \in (-\infty, 1]$, there exists a second-best policy that is Markov. That is, there exists $\underline{\mu}^{\text{SB}}$ such that in the second-best policy the designer recommends Y if $\mu > \underline{\mu}^{\text{SB}}$, N if $\mu < \underline{\mu}^{\text{SB}}$, and recommends Y with some probability $p \in [0, 1]$ if $\mu = \underline{\mu}^{\text{SB}}$.

Proof. First we consider the case $a > -1$. By Theorem 2, consumers' incentives are aligned with the designer if the designer has $a = -1$. As we will show below, instead of considering the original constrained optimization problem, it is easier to consider an equivalent problem with a constraint that the expected payoff of an auxiliary designer with $a = -1$ needs to be non-negative.

For a given recommendation policy r, the ex ante expected payoffs of the two designers are

$$V(r, \mu_0) = \sum_{t=1}^{\infty} \delta^{t-1} \mathbb{E}_{\pi,R}(v_t),$$
$$W(r, \mu_0) = \sum_{t=1}^{\infty} \delta^{t-1} \mathbb{E}_{\pi,R}(w_t).$$

Here, v_t and w_t are the stage-game payoffs at Period t for the actual designer ($a > -1$) and the auxiliary designer ($a = -1$), respectively. The expectation is taken over the signal structure π and the recommendation policy r, taken the prior belief μ_0 as given. By Proposition 3, either the first-best policy is implementable, or ICC-Y is binding while ICC-N is slack in any second-best policy. By ignoring ICC-N and writing ICC-Y as $W(R, \mu_0) \geqslant 0$, the designer's problem becomes:

$$\begin{aligned} \max_{r} \quad & V(r, \mu_0) \\ \text{subject to} \quad & W(r, \mu_0) \geqslant 0. \end{aligned} \tag{4}$$

Note that $W(r, \mu_0) \geqslant 0$ is ICC-Y, since if the auxiliary designer's ex ante expected payoff and the consumers' expected payoff of purchasing after recommendation Y share the same sign.

Since we have shown that the first-best policy is a threshold policy in Theorem 1, we will focus on the case that the first-best policy is not implementable. That implies $W(r^{\mathrm{SB}}, \mu_0) = 0$ under the second-best policy r^{SB}. That is, the actual designer cannot improve his payoff without dropping the auxiliary designer's payoff below 0. Also note that the feasible set of (V, W) is convex because the designer can randomize between two policies at the beginning of the game. The arguments above allow us to use the weighting method[11]. Since $(V^{\mathrm{SB}}, 0)$ is Pareto optimal, there exists $\lambda \geqslant 0$ such that V^{SB} is the value of the non-constrained problem

$$\max_{r} \quad V(r, \mu_0) + \lambda W(r, \mu_0). \tag{5}$$

And since we assume that $W(r^{\mathrm{SB}}, \mu_0) = 0$, $\lambda > 0$. For the purpose of this proof, the value of λ is not important; it suffices to know that $\lambda > 0$.

(5) is a scalarized problem. Since the stage payoffs depend on belief only, the signal depends on the true quality state only, and the belief transition is Markov, (5) is a standard Markov decision problem that can be solved by using the Bellman equation:

$$V_\lambda(\mu) = \max\left\{(1-a)\mu + a + \lambda(2\mu - 1) + \delta\mathbb{E}_\pi[V_\lambda(\mu') \mid \mu], \quad \delta V_\lambda(\mu)\right\}. \tag{6}$$

Since $\lambda > 0$, the stage game payoff $(1-a)\mu + a + \lambda(2\mu - 1)$ is strictly increasing in belief μ. This means the solution of (6) admits a threshold policy by applying a similar argument as in the proof of Theorem 1. That is, regarding (6), there exists a threshold $\underline{\mu}^{\mathrm{SB}}$ such that an experiment is preferred if $\mu > \underline{\mu}^{\mathrm{SB}}$, indifferent if $\mu = \underline{\mu}^{\mathrm{SB}}$, and not preferred if $\mu < \underline{\mu}^{\mathrm{SB}}$. This defines a family of experimentation policies $\tilde{r}(p)$ parameterized by a number $p \in [0, 1]$, where p is the probability of experimenting at belief $\underline{\mu}^{\mathrm{SB}}$.

We argue that there exists p such that $\tilde{r}(p)$ is a second-best policy of the problem in (4) (hence a second-best policy of the original problem). Since the payoff difference between experimenting or not,

$$(1-a)\mu + a + \lambda(2\mu - 1) + \delta\mathbb{E}_\pi[V_\lambda(\mu') \mid \mu]$$

[11] See, for example, Theorem 3.1.4 in Miettinen (1999), p.79.

is strictly increasing in μ, $\tilde{r}(p)$ is a complete Bellman characterization, in the sense that every Bellman-optimal policy is of the form of $\tilde{r}(p)$. Since $(V^{\text{SB}}, 0)$ is an optimal value pair, there must exist $p^{\text{SB}} \in [0,1]$ such that $V(\tilde{r}(p^{\text{SB}})) = V^{\text{SB}}$ and $W(\tilde{r}(p^{\text{SB}})) = 0$. Therefore, $\tilde{r}(p^{\text{SB}})$ is a second-best policy that experiments at all beliefs higher than $\underline{\mu}^{\text{SB}}$, experiments with probability p^{SB} at $\underline{\mu}^{\text{SB}}$, and does not experiments at beliefs lower than $\underline{\mu}^{\text{SB}}$. When $p^{\text{SB}} \notin \{0,1\}$, the second-best policy involves mixing at the threshold belief $\underline{\mu}^{\text{SB}}$.

We now consider the case $a < -1$. Again, it suffices to consider cases when the first best is not implementable. By Proposition 3, ICC-Y is slack while ICC-N binds in the second best. Similar to above, we can ignore ICC-Y for now and formulate ICC-N with an auxiliary designer whose a is -1. The stage payoffs of the auxiliary designer needs to be different though: the auxiliary designer gets the payoffs from experiments when N is recommended and 0 when Y is recommended. Therefore, the ex ante payoff of this auxiliary designer, denoted by $\bar{W}$, is the *deviation* payoff after observing N. The transformed problem is now

$$\begin{aligned} \max_{r} \quad & V(r, \mu_0) \\ \text{subject to} \quad & \bar{W}(r, \mu_0) \leqslant 0. \end{aligned} \tag{7}$$

To use the weighting method on (7), we need one more step: write ICC-N as $-\bar{W}(r, \mu_0) \geqslant 0$, and the second best is a point on the Pareto frontier of $(V, -\bar{W})$. The non-constrained problem is

$$\max_{r} \quad V(r, \mu_0) + \bar{\lambda}[-\bar{W}(r, \mu_0)]. \tag{8}$$

By adding the negative sign, we guarantee that $\bar{\lambda} \geqslant 0$. Since ICC-N must bind, $\bar{\lambda} > 0$. The scalarized problem (8) again satisfies the standard Markov conditions and can be formulated using Bellman equation. Notice that the stage payoff is $\mu + (1-\mu)a = a + (1-a)\mu$ when Y is recommended by the recommendation policy, and the stage payoff is $\bar{\lambda}(1-2\mu)$ when N is recommended. Therefore, the Bellman equation is

$$\bar{V}_{\bar{\lambda}}(\mu) = \max\{a + (1-a)\mu + \delta \mathbb{E}_{\pi}[\bar{V}_{\bar{\lambda}}(\mu') \mid \mu, Y], \quad \bar{\lambda}(1-2\mu) + \delta \mathbb{E}_{\pi}[\bar{V}_{\bar{\lambda}}(\mu') \mid \mu, N]\}.$$

When there is no experiment, the belief remains the same, so that $\mathbb{E}_{\pi}[\bar{V}_{\bar{\lambda}}(\mu') \mid \mu, N] = \bar{V}_{\bar{\lambda}}(\mu)$. Therefore, at a belief where N is recommended,

$$\bar{V}_{\bar{\lambda}}(\mu) = \bar{\lambda}(1-2\mu) + \delta \bar{V}_{\bar{\lambda}}(\mu),$$

which is

$$\bar{V}_{\bar{\lambda}}(\mu) = \frac{\bar{\lambda}(1-2\mu)}{1-\delta},$$

so that the Bellman equation can be written as

$$\bar{V}_{\bar{\lambda}}(\mu) = \max\left\{a + (1-a)\mu + \delta \mathbb{E}_{\pi}[\bar{V}_{\bar{\lambda}}(\mu') \mid \mu, Y], \quad \frac{\bar{\lambda}(1-2\mu)}{1-\delta}\right\}. \tag{9}$$

To show that a second-best policy takes a threshold form, we need to show that the difference of the payoffs of recommending Y and the payoffs of recommending N increases as the belief increases. In other words, we need to show that

$$D_{\bar{\lambda}}(\mu) = a + (1-a)\mu + \delta \mathbb{E}_\pi[\bar{V}_{\bar{\lambda}}(\mu') \mid \mu, Y] - \frac{\bar{\lambda}(1-2\mu)}{1-\delta} \tag{10}$$

is strictly increasing in μ. This is less direct than the case with $a > -1$. In the latter cases, both the real designer and the auxiliary gets 0 from no experiment, and showing that V_λ is increasing is sufficient. Here, stage payoffs are not 0 in general. The real designer gets payoff from experiment, while the auxiliary designer gets payoff from no experiment. A direct result is that we lose the monotonicity of $\bar{V}_{\bar{\lambda}}$. This can be observed directly from (9): when N is played, $\bar{V}_{\bar{\lambda}} = \bar{\lambda}(1-2\mu)/(1-\delta)$ is decreasing in μ. This means that we cannot directly use the FOSD property to show the monotonicity of $D_{\bar{\lambda}}$. In what follows, we will introduce a transformation that shows the monotonicity of $D_{\bar{\lambda}}$. Define x^+ as a short-hand notation for $\max\{x, 0\}$. We can write (7) as

$$\bar{V}_{\bar{\lambda}}(\mu) = \frac{\bar{\lambda}(1-2\mu)}{1-\delta} + D_{\bar{\lambda}}(\mu)^+.$$

It follows that

$$\begin{aligned}
\mathbb{E}[\bar{V}_{\bar{\lambda}}(\mu') \mid \mu, Y] &= \mathbb{E}\left[\frac{\bar{\lambda}(1-2\mu')}{1-\delta} \;\middle|\; \mu, Y\right] + \mathbb{E}[D_{\bar{\lambda}}(\mu')^+ \mid \mu, Y] \\
&= \frac{\bar{\lambda}(1-2\mu)}{1-\delta} + \mathbb{E}[D_{\bar{\lambda}}(\mu')^+ \mid \mu, Y],
\end{aligned}$$

where the second equality follows from the law of iterated expectations. Substitute this back to (10),

$$D_{\bar{\lambda}}(\mu) = a + (1-a)\mu + \mathbb{E}[D_{\bar{\lambda}}(\mu')^+ \mid \mu, Y].$$

The goal of this transformation is to write $D_{\bar{\lambda}}(\mu)$ in a recursive form to be able to use standard techniques involving value function iteration. Indeed, using a similar argument as in Theorem 1, define

$$(\bar{T}f)(\mu) = a + (1-a)\mu + \mathbb{E}[f(\mu')^+ \mid \mu, Y].$$

If f is an increasing function, $f^+ = \max\{f(\mu), 0\}$ is also increasing. By the FOSD property (implied by the MLRP property) of the signal structure, $\mathbb{E}[f(\mu)^+ \mid \mu, Y]$ is increasing in μ. Since $a < -1$, the first two terms $a + (1-a)\mu$ is strictly increasing in μ. Therefore, $\bar{T}$ preserves the monotonicity. The standard value function iteration argument then indicates that $D_{\bar{\lambda}}$ is strictly increasing in μ. This implies the threshold structure of the second-best policy. What remains is to establish the possibility of mixing at the threshold, which is similar to the case $a > -1$ and omitted here. □

The threshold belief $\underline{\mu}_{\text{SB}}$ in general depends on the prior belief μ_0, the designer's experimentation incentives a, the arrival probability δ, and the informativeness of the feedback signal structure π.

The proof of Theorem 4 uses the observation in Theorem 2 that for any given recommendation policy, the ex ante payoff of a designer with $a = -1$ has the same sign as the expected payoff of a time-blind consumer following a purchase recommendation. When $a > -1$, ICC-Y is the relevant binding constraint. We therefore introduce an auxiliary designer with $a = -1$ and use his ex ante payoff to represent ICC-Y. Because the actual and auxiliary designers' payoffs have the same intertemporal structure and differ only in their flow payoffs, the constrained problem can be scalarized by assigning a weight to the auxiliary designer's payoff. This converts the problem into a standard recursive Markov decision problem.

When $a < -1$, the relevant binding constraint is instead ICC-N. In this case, we construct the auxiliary payoff so that it records the consumer's deviation payoff when N is recommended and assigns zero payoff when Y is recommended. After incorporating the negative of this payoff into the actual designer's objective, the problem can again be scalarized and written recursively. In both cases, the scalarized dynamic program has a monotone action-value difference, which yields the threshold structure, while the multiplier associated with the binding IC constraint pins down the particular second-best threshold.

Theorem 4 shows that for general learning signal structures the second-best policies for the designer can be restricted to belief threshold policies without loss of generality. How this shapes optimal recommendation policies and how much influence the designer can have on experimentation depend on the underlying learning signal structures, as we demonstrate in the next sections.

5 The Designer's Second Best for Perfect Good News

Consider the signal structure of perfect good news from Example 1, but now in the infinite horizon framework of our base model. Given this signal structure, there are only countably many possible beliefs on path, and the belief monotonically decreases when no good news arrives after a purchase. This also allows us to abstract away from specific histories, as the belief itself contains the relevant experimentation information. Let μ_n be the belief at the beginning of a period if there were $n-1$ purchases before but no good news having arrived. Let $\mu_1 = \mu_0$ be the belief at the beginning of Period 1.

The monotonically decreasing belief in the absence of good news implies a simple form of the second-best policy. Following from Proposition 3 and Theorem 4, the designer's second-best policy involves two phases: an *exploration phase*, where the designer experiments at the beginning of the game, and an *exploitation phase*, where the designer only recommends purchase if news arrived and the product was confirmed to be good. The threshold belief is chosen according to the first best (when a is around -1) or a binding IC constraint (when a is significantly larger or smaller than -1), and mixing at the end of the exploration phase is in general necessary in the latter case.

To provide a full characterization of the designer's second-best policy, we employ the occupancy measure approach again, with relatively straightforward definitions of occupancy measures due to the simple belief evolution:

- Define x_n as the expected number of periods when there is a consumer present and the designer's belief at the beginning of the period is μ_n.
- Define y_n as the expected number of purchases made when the belief is μ_n.

This way recommendations at histories with a given belief are implicitly reflected by the relation between x_n and y_n. In particular, if $y_n = x_n$, we know that Y is recommended whenever the belief is μ_n. If $0 < y_n < x_n$ then at some periods where the belief is μ_n and there is a consumer, N is recommended. If $y_n = 0$ then N is always recommended at belief μ_n. By definition,

$$\begin{aligned} x_1 &= 1 + \delta(x_1 - y_1), \\ x_{n+1} &= \delta(1 - \alpha\mu_n)y_n + \delta(x_{n+1} - y_{n+1}), \quad n > 1. \end{aligned} \tag{11}$$

Consider $n > 1$. The first term on the RHS is the inflow of x_{n+1}: it counts the expected number of periods x_{n+1} is reached via experimentation. The second term considers if the designer mixes and the realization is recommending N, so that the belief at the beginning of the next period is still μ_{n+1}. For $n = 1$, since the first consumer arrives for sure, and $\mu_1 = \mu_0$, the inflow term is 1. For concreteness, we show the law of motion for $n = 1$. Let ρ_t be the probability of recommending Y to consumer t assuming that the belief at the beginning of Period t is μ_1. Then,

$$\begin{aligned} x_1 &= 1 + (1 - \rho_1)\delta + (1 - \rho_1)(1 - \rho_2)\delta^2 + (1 - \rho_1)(1 - \rho_2)(1 - \rho_3)\delta^3 + \cdots \\ y_1 &= \rho_1 + (1 - \rho_1)\delta\rho_2 + (1 - \rho_1)(1 - \rho_2)\delta^2\rho_3 + \cdots \end{aligned}$$

For x_1, the t-th term on the RHS is the probability that the belief at the beginning of Period t is μ_1. Since the first consumer arrives for sure, this probability is 1 for $t = 1$. For Period 2, the belief is μ_1 if the designer recommends N (note that an experiment will change the belief for sure), which happens with probability $1 - \rho_1$, and a consumer arrives with probability δ. y_1 is formulated in the similar way, with the t-th term on the RHS being the probability of recommending Y at Period t under belief μ_0. It is easy to see that

$$x_1 - y_1 = 1 - \rho_1 + (1 - \rho_1)(1 - \rho_2)\delta + \cdots ,$$

hence $x_1 = 1 + \delta(x_1 - y_1)$.

With perfect good news, the arrival of good news implies that the belief jumps up to 1. Let x_G be the expected number of periods staying at belief 1 and with a consumer arrival. Then,

$$x_G = \delta x_G + \delta \sum_{n \geqslant 1} \alpha\mu_n y_n.$$

This gives

$$x_G = \frac{\delta}{1-\delta} \sum_{n \geqslant 1} \alpha \mu_n y_n.$$

The designer always recommends purchasing at belief 1 in any second-best policy, since increasing the probability of recommending purchasing at belief 1 both strictly increases the designer's objective function and relaxes the corresponding IC constraint.

The use of occupancy measures simplifies the designer's problem. As we will show next, the designer's problem reduces to a linear programming problem, and the designer's second-best policies are unique when described by occupancy measures.

5.1 The Designer's Problem

We start with the IC constraints. The expected belief after recommendation Y is

$$\frac{\sum_{n \geqslant 1} \mu_n y_n + 1 \cdot x_G}{\sum_{n \geqslant 1} y_n + x_G}$$

A consumer follows the recommendation Y if this is at least 1/2. Plugging in x_G ICC-Y can be rewritten as

$$\sum_{n \geqslant 1} \left[\mu_n \left(2 + \frac{\alpha \delta}{1-\delta} \right) - 1 \right] y_n \geqslant 0. \tag{12}$$

The expected belief after recommendation N is

$$\frac{\sum_{n \geqslant 1} (x_n - y_n) \mu_n}{\sum_{n \geqslant 1} (x_n - y_n)},$$

and this needs to be less than 1/2 for incentive compatibility, which corresponds to

$$\begin{aligned}
0 &\geqslant \sum_{n \geqslant 1} (x_n - y_n)(2\mu_n - 1) \\
&= (2\mu_1 - 1)(x_1 - y_1) + \sum_{n \geqslant 2} (x_n - y_n)(2\mu_n - 1) \\
&= (2\mu_1 - 1) \left(\frac{1 - \delta y_1}{1-\delta} - y_1 \right) + \sum_{n \geqslant 2} (2\mu_n - 1) \left(\frac{\delta}{1-\delta} [(1 - \alpha \mu_{n-1}) y_{n-1} - y_n] - y_n \right),
\end{aligned}$$

where the last line uses (11). Multiplying both sides by $(1-\delta)$ and collect terms, we get

$$\begin{aligned}
0 &\geqslant 2\mu_1 - 1 + \sum_{n \geqslant 1} [\delta (1 - \alpha \mu_n)(2 - \mu_n - 1) - (2\mu_{n+1} - 1)] y_n \\
&= 2\mu_1 - 1 + \sum_{n \geqslant 1} \big(1 - \delta - [2(1-\delta) + \alpha \delta] \mu_n \big) y_n,
\end{aligned} \tag{13}$$

where the last line uses $\mu_{n+1} = \mu_n (1-\alpha)/(1 - \alpha \mu_n)$.

Equation (13) is the ICC-N. Since $\{\mu_n\}$ is a decreasing sequence, the coefficient of y_n is increasing in n, so a positive y_n tightens ICC-N when n is sufficiently large. A positive y_n at such μ_n means recommending Y with positive probability, decreasing the probability of recommending N at that belief. In turn, both IC constraints become harder to satisfy.

The objective function of the designer is also straightforward to express using x_G and y_n:

$$\sum_{n\geqslant 1}[(1-a)\mu_n + a]y_n + x_G,$$

which simplifies to

$$\sum_{n\geqslant 1}\left[\mu_n\left(1-a+\frac{\alpha\delta}{1-\delta}\right)+a\right]y_n. \tag{14}$$

The designer's problem is to choose $\{x_n, y_n\}$ to maximize (14), subject to the feasibility constraint $0 \leqslant y_n \leqslant x_n$, the law of motion (11), and the IC constraints (12) and (13).

Observe that when $a = -1$, the objective function (14) coincides with the LHS of the IC constraint (12). This verifies that the first-best policy can be implemented when the designer and consumers have aligned preferences.

5.2 The second-best policies

By Proposition 3 and Theorem 4, there exists some n such that $y_k = x_k$ for $k < n$, $y_k = 0$ for $k > n$. In other words, the designer will front-load experimentation to the beginning of the game. Intuitively, since the designer and consumers weigh good and bad products differently, y_m and y_{m+k}, when adjusted by the IC constraint, do not contribute to the designer's objective function in the same way, and the designer should front-load the experimentation, to histories where the belief is higher. This shows that the mixing can only happen at the end of the exploration phase. We only need to determine the value of n, which pins down the threshold belief, and the value of y_n, which determines the mixing behavior at the threshold. When the second best involves a binding IC constraint, the second-best policy is also unique with respect to the occupancy measure.

We next offer a characterization of the designer's second-best policy. For simpler notations later, let $\underline{\mu}(0) = 0$, $w(0) = 1/2$, $s(0) = 0$, and for $n = 1, 2, \cdots$, define

$$\begin{aligned}
\underline{\mu}(n) &= \frac{-a\delta^{n-1}(1-\delta)}{-a\delta^{n-1}(1-\delta) + [1-\delta(1-\alpha)]\,[\delta(1-\alpha)]^{n-1}},\\
w(n) &= \frac{1}{(1-\alpha)^n + 1},\\
s(n) &= \frac{1-\delta^n}{2-[\delta(1-\alpha)]^n - \delta^n}.
\end{aligned}$$

Here, as we will show later, $\underline{\mu}(n)$ is the lowest belief at which the designer would like to experiment n times without arrivals of good news. In particular, $\underline{\mu}(1) = \underline{\mu}$, the first-best boundary to start any experimentation. $w(n)$ is the boundary determined by ICC-N. If the prior is $w(n)$, the posterior belief after n experiment is exactly $1/2$. In other words, the recommendation policy must recommend Y for at least n times before recommending N forever. $s(n)$ is the boundary determined by ICC-Y, which is the lowest prior belief at which the consumer is willing to accept Y from a policy that experiments exactly n times.

In the following theorems in this section, we characterize the designer's second-best policies by $\{\underline{\mu}(n), w(n), s(n)\}_{n=0}^{\infty}$. We illustrate the designer's second-best policies in Figure 1. In general, the designer is more willing to initiate experimentation when a is higher (reflecting stronger experimentation incentives), when α is higher (as more precise experiments improve the information values of experimentation), or when δ is higher (implying a larger expected number of consumers). Consumers are also more willing to accept the recommendation when α is higher, since greater accuracy enables the designer to provide better recommendations, or when δ is higher, as a consumer believes it with higher probability that she arrives after perfect news already having arrived. We start from a designer who prefers less experimentation, namely $a < -1$.

Theorem 5. Let $a < -1$. For $\underline{\mu}(n) \leqslant \mu_0 < \underline{\mu}(n+1)$, $n = 0, 1, \cdots$, there exists an optimal policy as following:

- If $\mu_0 \leqslant w(n)$, implement the first-best policy. That is, recommend Y until the n-th period, then recommend N forever, if there is no good news arriving. After an arrival of good news, recommend Y forever.
- If $\mu_0 > w(n)$, recommend Y until the $(n+k)$-th period, mix at the $(n+k+1)$-th period, and then recommend N forever, where k satisfies $w(n+k) < \mu_0 \leqslant w(n+k+1)$, if there is no good news arriving. After an arrival of good news, recommend Y forever.

Proof. By induction,

$$\mu_n = \frac{\mu_0(1-\alpha)^{n-1}}{\mu_0(1-\alpha)^{n-1} + 1 - \mu_1},$$

and if $y_j = x_j$ for all $j \leqslant n$,

$$y_n = \delta^{n-1}[\mu_0(1-\alpha)^{n-1} + 1 - \mu_1].$$

When $a < -1$, only ICC-N (13) potentially binds. Substituting μ_n and y_n into the RHS of the ICC-N, the RHS becomes

$$\mu_0([\delta(1-\alpha)]^n + \delta^n) - \delta^n.$$

Set this to 0 to get $\mu_0 = w(n)$. That is, ICC-N binds if the prior is $w(n)$ and the designer experiments exactly n times. One can verify that, by substituting $w(n)$ into the equation of μ_n, $\mu_{n+1} \mid_{\mu_0 = w(n)} = 1/2$. That is, the designer will experiment as long as the belief at the beginning of the period is above $1/2$, otherwise ICC-N gets violated.

When $\underline{\mu}(n) \leqslant \mu_0 < \underline{\mu}(n+1)$, the designer would like to experiment n times in the first best. If $\mu_0 \leqslant w(n)$, this means the belief after n experiments without good news is at most $1/2$, suggesting that the first best is implementable. Otherwise, the first best is not implementable, and the second best is determined by ICC-N. Note that $w(k) \to 1$ as $k \to \infty$, so for any $\mu_0 \in (0,1)$, there exists $k \geqslant 0$ such that $w(n+k) < \mu_0 \leqslant w(n+k+1)$. For $w(n+k) < \mu_0 \leqslant w(n+k+1)$, the designer must experiment at the first $n+k$ beliefs and mix at the belief $n+k+1$ to make the policy just implementable. By letting $y_j = x_j$ for $j \leqslant n+k$, setting ICC-N to be 0 gives

$$y_{n+k+1} = \left[\delta^{n+k} - \mu_0\left([\delta(1-\alpha)]^{n+k} + \delta^{n+k}\right)\right] \Big/ \left[1 - \delta - [2(1-\delta) + \alpha\delta]\frac{\mu_1(1-\alpha)^{n+k}}{\mu_1(1-\alpha)^{n+k} + 1 - \mu_1}\right].$$

□

When $a < -1$, only ICC-N matters by Proposition 3, so the constraint comes from making recommendations of N credible. To ensure this, if no good news arrives, the designer must terminate at a belief of $1/2$ or below. When a is near -1, the first best terminates at a belief of $1/2$ or lower and thus is implementable. As a becomes smaller, the designer's incentive to experiment eventually becomes too weak so that terminating at such a belief is not his first best. In such cases, the designer has to experiment more than his first best until the termination belief is exactly $1/2$, which typically requires mixing at the end of the experimentation.

Next, consider a designer who prefers more experimentation but would not want to recommend a known bad product.

Theorem 6. Let $-1 < a < 0$. For $\underline{\mu}(n) \leqslant \mu_0 < \underline{\mu}(n+1)$, $n = 0, 1, \cdots$, there exists an optimal policy as following:

- If $\mu_0 \geqslant s(n)$, implement the first-best policy. That is, recommend Y until the n-th period, then recommend N forever, if no good news is arriving.
- If $s(k) \leqslant \mu_0 < s(k+1)$ for $k = 1, \cdots, n-1$, recommend Y until the k-th period, mix at the $(k+1)$-th period, then recommend N forever, if no good news is arriving.
- If $\mu_0 < s(1)$, recommend N forever.

Proof. When $-1 < a < 0$, only ICC-Y (12) potentially binds. Substituting μ_n and y_n in the proof of Theorem 5 into the LHS of the ICC-Y yields

$$\mu_0 \frac{2 - [\delta(1-\alpha)]^n - \delta^n}{1-\delta} - \frac{1-\delta^n}{1-\delta}.$$

Set this to 0 to get $\mu_0 = s(n)$. That is, if $\mu_0 = s(n)$, under a policy in which the designer experiments exactly n times without good news, the expected payoff of purchasing the product after a recommendation Y is exactly 0.

If $\underline{\mu}(n) \leqslant \mu_0 < \underline{\mu}(n+1)$, the designer again would like to experiment n times in the first best. This is implementable if $\mu_0 \geqslant s(n)$ by construction. Otherwise, the belief is too low to implement the first best. In particular, for $\mu_0 < s(1)$, the consumers cannot accept any recommendation Y. For $\mu_0 < s(n)$ and $n > 1$, there must exist a k such that $s(k) \leqslant \mu_0 < s(k+1)$ and $1 \leqslant k < n$. In this case, ICC-Y is slack after k experiments without good news arrival but is violated after $k+1$ experiments (by the construction of $s(n)$ again). The designer then uses mixing such that ICC-Y binds exactly. By letting $y_j = x_j$ for $j \leqslant k$ and setting the LHS of ICC-Y to be 0, we have

$$y_{k+1} = \left[\mu_0 \frac{2 - [\delta(1-k)]^k - \delta^k}{1-\delta} - \frac{1-\delta^k}{1-\delta}\right] \Big/ \left[1 - \frac{\mu_0(1-\alpha)^k}{\mu_0(1-\alpha)^k + 1 - \mu_0} \frac{2 - 2\delta + \alpha\delta}{1-\delta}\right].$$

□

When $-1 < a < 0$, only ICC-Y matters, so the constraint comes from making recommendations of Y credible. The designer's first best is always implementable when $\mu_0 > 1/2$ because the consumers are willing to follow a recommendation of Y even without information, and the designer's experimentation incentive is even stronger than the consumers'. Note that this statement is also true when $a \geqslant 0$. When $\mu_0 < 1/2$, the designer can implement his first best when a is near -1. However, he has to implement a policy different from his first best either when a becomes high enough such that he wants experimentation more than ICC-Y would have allowed, or when μ_0 becomes low enough such that he has to experiment less to make a recommendation of Y credible.

Finally, consider a designer who has a strong experimentation incentive and would even want to recommend a known bad product.

Theorem 7. Let $a \geqslant 0$. There exists an optimal policy as following:

- If $\mu_0 \geqslant 1/2$, implement the first-best policy. That is, recommend Y forever.
- If $\mu_0 < s(1)$, recommend N forever. If $s(k) \leqslant \mu_0 < s(k+1)$ for some $k \geqslant 1$, recommend Y until the k-th period, mix at the $(k+1)$-th period, then recommend N forever.

Proof. When $a \geqslant 0$, the designer would like to recommend Y to all consumers, which is only IC if $\mu_0 \geqslant 1/2$. For $\mu_0 < 1/2$, ICC-Y binds, and the second-best policy is pinned down by ICC-Y. Again, there exists $k \geqslant 1$ such that $s(k) \leqslant \mu_0 < s(k+1)$. In this case, the designer can at most experiment k times with probability 1 while keeping ICC-Y. y_{k+1} is again solved and equals the result in Theorem 6 by letting $y_j = x_j$ for $j \leqslant k$ and setting the LHS of ICC-Y to be 0. □

Given that $a \geqslant 0$, the designer's first best is to recommend Y forever, which is implementable only when $\mu_0 > 1/2$. When $\mu_0 < 1/2$, the best the designer can achieve is to recommend the product as many periods as ICC-Y allows, which is determined by $\{s(n)\}_{n=0}^{\infty}$.

We now summarize the designer's second-best policies for different μ_0 and a in Figure 1. As the color shifts from gray to orange, the designer gradually increases the periods of experimentation.

In the regions shaded with solid colors, the designer's first-best policies are implemented. Also, the first-best policies are always implementable in the region of $\mu_0 \geqslant 1/2$ and $a > -1$. In the regions shaded with vertical patterns, the first best cannot be implemented, and the designer in general employs mixing at the end of the experimentation phase. In such regions, the occupancy measure corresponding to mixing increases as μ increases but does not depend on a, since it is determined by the IC constraints only.

Even though the second-best policy is unique when one IC constraint binds, there are multiple ways to implement the same occupancy measure if the designer's second-best policy involves a (non-degenerate) mixed recommendation at μ_n. In Appendix A, we provide an example that shows the designer can delay experimentation at a belief if the recommendation policy uses mixing at this belief. In general, when the second best involves mixing before recommending N for sure, the designer can choose to front-load the experiment by mixing only once at the first period when μ_n is reached, or delay and spread the experiment for several periods while the belief is still at μ_n. Notice that this does not contradict the observation that the designer front-loads the experimentation, since the delayed experimentation happens at the same belief. In other words, while the belief is μ_n, two recommendation policies yield the same designer's payoff and keep the same incentive compatibility as long as the expected numbers of consumers who receive recommendations Y and N are the same. Also, delaying the experiment is only feasible when there is non-degenerate mixed recommendation. For belief μ_m where $y_m = x_m$, the designer cannot delay experiment, since there is a strictly positive probability that the consumers stop arriving in future periods.

6 The Designer's Second Best with Perfect Bad News

In this section, we investigate the designer's second-best policies when learning about the product takes the form of perfect bad news. More specifically, if the product is of bad quality, then following purchase the consumer (and through feedback the designer) receives bad news (b) with probability $\beta > 0$; otherwise there is no news (n). The signal structure is:

$$\begin{bmatrix} \mathbb{P}(b \mid G) & \mathbb{P}(n \mid G) \\ \mathbb{P}(b \mid B) & \mathbb{P}(n \mid B) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \beta & 1-\beta \end{bmatrix}.$$

The perfect bad news structure again yields a simple belief process: there are only countably many on-path beliefs, and the belief monotonically increases in the absence of news. This means the designer's first-best policy when $a < 0$ is either recommending Y until bad news arrives if $\mu_0 \geqslant \underline{\mu}$, or recommending N to all consumers if $\mu_0 < \underline{\mu}$, where $\underline{\mu}$ is the first-best threshold as defined in Theorem 1. However, an *increasing* belief sequence suggests potentially different mixing behavior when the first-best policy is not implementable. As suggested by Theorem 4, any mixing induced by IC constraints should occur at the lowest belief at which Y is recommended with

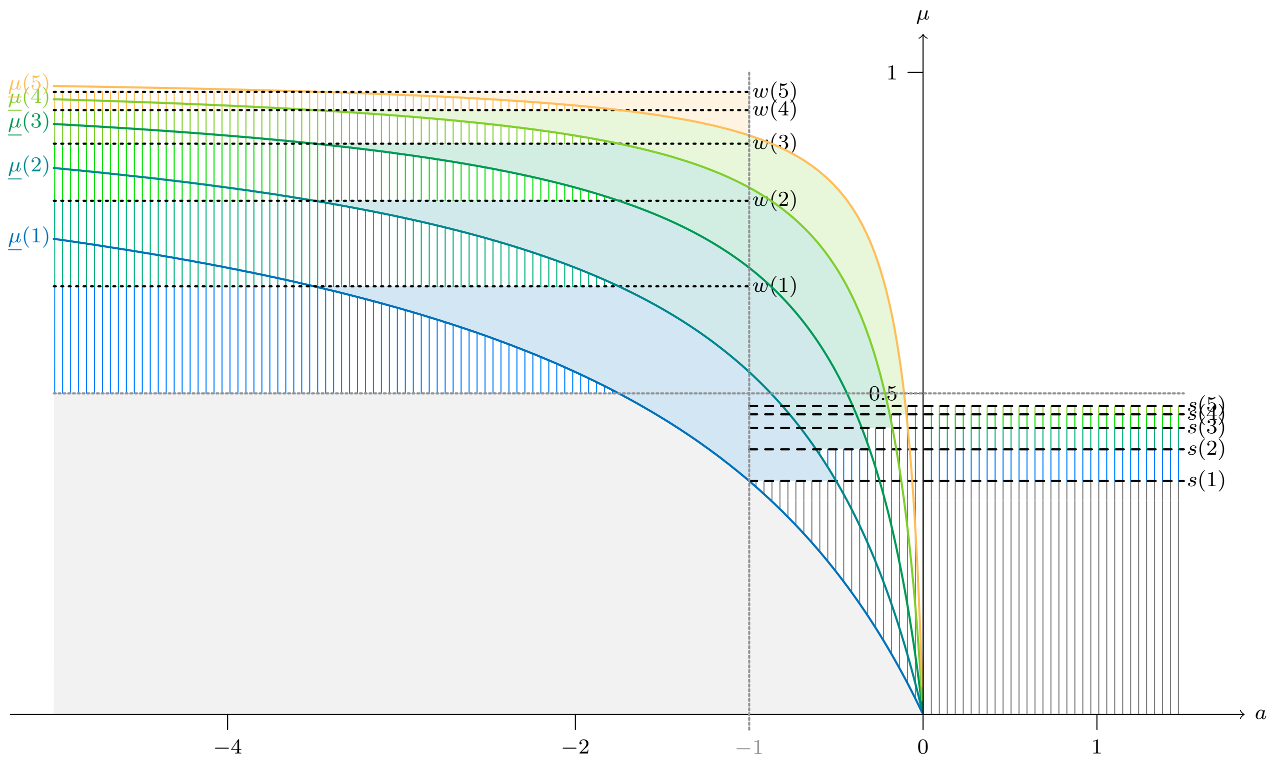

Figure 1 The second-best policy under perfect good news

Note: Drawn for $\alpha = 0.5$ and $\delta = 0.6$. For each pair (a, μ_0), the figure displays the number of rounds of experimentation conducted for sure under the second-best policy when no good news arrives.

- Regions with solid shading indicate parameter values for which the designer's first-best policy is incentive compatible and can therefore be implemented.
- Regions with vertical hatching indicate parameter values for which the first best is not incentive compatible. In these regions, the relevant IC constraint binds, and the designer generally mixes in the final round of the exploration phase.
- The shading ranges from gray to orange and represents 0 to 5 rounds of experimentation conducted for sure. In a hatched region, the designer may additionally experiment with positive probability in the following round.
- As defined before Theorem 5, the curves $\underline{\mu}(n)$ delimit the designer's first-best experimentation regions, while $s(n)$ and $w(n)$ delimit the regions determined by ICC-Y and ICC-N, respectively.
- For readability, the color scale is capped at 5 rounds. Regions corresponding to more than 5 rounds of experimentation, including the region in which purchase is always recommended, are left unshaded.

positive probability. Therefore, under perfect bad news where relatively low beliefs now occur at the beginning of the game (except following the arrival of bad news when the belief drops to 0), mixed recommendations, if any, should occur at the prior belief.

Nevertheless, mixing early may or may not help the designer depending on whether the designer has an over- or under-experimentation incentive. Intuitively, if $a < -1$, so the designer is less willing to recommend Y and can mix by recommending N with positive probability at the prior belief. Because consumers cannot distinguish between such N recommendations and an N recommendation following bad news, this mixing can increase the designer's payoff. By contrast, when $a > -1$, early mixing may not benefit the designer. The consumers now worry about over-experimentation, and the goal of mixing at the prior belief is to introduce uncertainty which reduces the proportion of Y recommendations at lower beliefs. But an observed recommendation Y resolves the uncertainty regarding the outcome of mixing and allows the consumers to condition on the realized recommendation being Y. This renders the mixing ineffective and leads to a second best that is either the first best or no experimentation at all. In what follows, we again use occupancy measures to formalize this intuition.

6.1 The Designer's Problem

We again let x_n denote the expected number of periods in which a consumer arrives and the belief at the beginning of the period is μ_n, where μ_n is the belief after $n-1$ purchases without bad news. Let y_n be the expected number of periods in which Y is recommended at belief μ_n. Similarly, define x_B as the number of periods in which the belief is 0 and a consumer arrives, and y_B as the number of periods in which a belief-0 product is recommended. The law of motion for the occupancy measure can be formulated similarly to subsection 5.1:

$$\begin{aligned} x_1 &= 1 + \delta\,(x_1 - y_1)\,, \\ x_{n+1} &= \delta\, y_n\,[1-(1-\mu_n)\,\beta] + \delta\,(x_{n+1} - y_{n+1})\,, \quad n \geqslant 1, \\ x_B &= \delta x_B + \delta \sum_{n\geqslant 1} y_n\,(1-\mu_n)\,\beta. \end{aligned} \tag{15}$$

ICC-N requires that consumers' expected belief after observing N is at most $1/2$, that is,

$$\frac{\sum_{n\geqslant 1}\mu_n(x_n - y_n) + (x_B - y_B)\cdot 0}{\sum_{n\geqslant 1}(x_n - y_n) + x_B - y_B} \leqslant \frac{1}{2}.$$

Equivalently,

$$\sum_{n\geqslant 1}(2\mu_n - 1)(x_n - y_n) - (x_B - y_B) \leqslant 0.$$

By Proposition 3, this IC binds only if $a < -1$, and in that case $y_B = 0$. Substituting for x_n and

x_B using (15), setting $y_B = 0$, and collecting terms by y_n yields

$$2\mu_1 - 1 + \sum_{n\geqslant 1} \{1 - 2\mu_n - \delta\beta(1-\mu_n) + \delta[1-(1-\mu_n)\beta](2\mu_{n+1}-1)\}\, y_n \leqslant 0.$$

Since $\mu_{n+1} = \mu_n/[1-(1-\mu_n)\beta]$, this simplifies to

$$2\mu_1 - 1 + \sum_{n\geqslant 1}(1-\delta)(1-2\mu_n)y_n \leqslant 0, \tag{16}$$

which is ICC-N under perfect bad news.

ICC-Y is similarly written as

$$\frac{\sum_{n\geqslant 1}\mu_n y_n + 0\cdot y_B}{\sum_{n\geqslant 1} y_n + y_B} \geqslant \frac{1}{2}.$$

Unlike ICC-N, this IC binds only if $a > -1$. If $a \geqslant 0$, the designer has incentives to recommend Y even when the product has belief 0, i.e., is known to be bad. Thus, we cannot in general set $y_B = 0$. The constraint can be rewritten as

$$\sum_{n\geqslant 1}(2\mu_n - 1)y_n - y_B \geqslant 0. \tag{17}$$

The designer's objective function is

$$\sum_{n\geqslant 1}[(1-a)\mu_n + a]y_n + a y_B. \tag{18}$$

The designer's problem is to maximize (18) subject to the feasibility constraints $0 \leqslant y_n \leqslant x_n$, $0 \leqslant y_B \leqslant x_B$, the law of motion (15), and IC constraints (16) or (17) (only one of them being relevant, as shown earlier), depending on the value of a.

6.2 The second-best policies

As noted above, unlike the perfect good news case, mixing need not appear in the second-best policy depending on the value of a. In this subsection, we consider the cases of $a < -1$ and $a > -1$ separately.

6.2.1 Designers with Weaker Experimentation Incentives

We start with the case $a < -1$, where ICC-N (16) is the relevant IC and $y_B = 0$. The designer's problem is again a linear programming problem, the solution of which determines the value of y_1. By Theorem 4, if $y_1 > 0$, $y_n = x_n$ for all $n > 1$. If $y_1 = 0$, then no experimentation is possible.

Theorem 8. Let $a < -1$. There exists a designer's second-best policy that

- recommends Y until bad news arrives if $\mu_0 \geqslant \underline{\mu}$.
- recommends N to all consumers if $\mu_0 \leqslant \min\{\underline{\mu}, 1/2\}$.
- a policy corresponding to

$$y_1 = \frac{(2\mu_0 - 1)[1 - (1-\beta)\delta]}{1 - \delta - 2\mu_0 + (2-\beta)\delta\mu_0}$$

  and $y_n = x_n$ for all $n \geqslant 2$, if $1/2 < \mu_0 < \underline{\mu}$.

Proof. First suppose that $\mu_0 \geqslant \underline{\mu}$. Since $a < -1$, ICC-Y is satisfied under the first-best policy (which is recommending Y until bad news arrives). Since N is recommended only for known bad products, ICC-N is also satisfied. Therefore, the first-best policy is implemented as the second-best policy when $\mu_0 \geqslant \underline{\mu}$. This is the first bullet point.

When $\mu_0 < \underline{\mu}$, the designer's first-best policy is to recommend N to all consumers, which is implementable only if $\mu_0 \leqslant 1/2$. This is the second bullet point.

For the remaining cases, by Theorem 4, the designer needs to choose y_1 so that ICC-N binds. Given that $y_n = x_n$ for $n \geqslant 2$, ICC-N is

$$(2\mu_0 - 1)(x_1 - y_1) - x_B = 0.$$

From (15),

$$x_1 = \frac{1 - \delta y_1}{1 - \delta},$$

and we need an expression of $x_n(1-\mu_n)\beta$ to write down x_B. It can be shown by induction that

$$x_n(1-\mu_n)\beta = \delta^{n-1} y_1 (1-\mu_0)(1-\beta)^{n-1}\beta,$$

so that

$$x_B = \frac{\delta y_1 \beta(1-\mu_0)}{(1-\delta)[1-\delta(1-\beta)]}.$$

Substituting x_1 and x_B into ICC-N, we get y_1 as stated. □

By Theorem 4, the designer's second-best policy is a threshold policy, where mixing only happens at the lowest belief with positive recommendation probability. Under a perfect bad news information structure, such a belief, if any, can only be the prior. To make recommendations of N credible, in any case where the designer's first best is not implementable, the designer mixes at the prior such that a consumer cannot tell whether an N comes from mixing or after conclusive bad news.

6.2.2 Designers with Stronger Experimentation Incentives

Consider now $a > -1$. The relevant IC constraint is ICC-Y (17). Again, Theorem 4, in the context of perfect bad news, suggests that mixed recommendations, if any, can only appear at the prior belief. We now formalize the intuition that mixing is ineffective and show that the designer will not even mix at μ_0.

Proposition 9. Suppose that $a > -1$. In any second-best policy, $y_1 = 0$ or $y_1 = 1$.

Proof. By Theorem 4 and (15), for $n \geqslant 1$,

$$y_{n+1} = \delta[1 - (1 - \mu_n)\beta]y_n.$$

In particular, this shows y_1 is a common factor of all y_n for $n \geqslant 2$. We can now write ICC-Y as

$$-y_B + y_1 \sum_{n \geqslant 1} (2\mu_n - 1)\tilde{y}_n \geqslant 0,$$

where

$$\tilde{y}_1 = 1, \quad \tilde{y}_2 = \frac{y_2}{y_1} = \delta[1 - (1 - \mu_1)\beta], \quad \text{and} \quad \tilde{y}_{n+1} = \frac{y_n}{y_1} = \delta[1 - (1 - \mu_1)\beta]\tilde{y}_n,$$

for $n \geqslant 2$. Observe that $\tilde{y}_n$ does not depend on y_1 for each n. Instead, it is determined by exogenous parameters μ_0, δ, and β. Therefore, the sign of $\sum_{n \geqslant 1}(2\mu_n - 1)\tilde{y}_n$ is determined by the same exogenous parameters.

If $\sum_{n \geqslant 1}(2\mu_n - 1)\tilde{y}_n < 0$, the only way to satisfy ICC-Y is to set $y_1 = 0$ and $y_B = 0$. If $\sum_{n \geqslant 1}(2\mu_n - 1)\tilde{y}_n \geqslant 0$, since $a > -1$,

$$\sum_{n \geqslant 1} [(1 - a)\mu_n + a]\tilde{y}_n > 0,$$

so that setting $y_1 = x_1$, the largest value possible, contributes to higher payoffs for both the designer and consumers who receive recommendation Y. By (15), if $y_1 = x_1$, then $y_1 = 1$. $\square$

When $a > -1$, only ICC-Y matters, and the constraint comes from making recommendations of Y credible. However, mixing at the prior belief cannot jump-start the experimentation if the prior is low. A consumer who receives Y correctly infers that the mixed recommendation must have realized as Y. Conditioning on this realization eliminates the benefit of mixing. Therefore, mixing cannot help increase the posterior belief after recommendation Y. For such low prior beliefs, the designer's only option is to recommend N forever.

We proceed by providing a characterization of the second-best policies using occupancy measures.

Theorem 10. Define

$$\mu^\star \equiv \frac{1-\delta}{2-2\delta+\beta\delta}.$$

For $-1 < a < 0$, there exists a designer's second-best policy that

- recommends Y until bad news if $\mu_0 \geqslant \mu^\star$. This is also the designer's first-best policy.
- recommends N to all consumers if $\mu_0 < \mu^\star$.

For $a \geqslant 0$, there exists a designer's second-best policy that

- recommends Y to all consumers if $\mu_0 \geqslant 1/2$. This is also the designer's first-best policy.
- recommends Y until bad news, and recommend known bad product with occupancy measure

$$y_B = \mu_0 \left[\frac{1}{1-\delta} + \frac{1}{1-(1-\beta)\delta}\right] - \frac{1}{1-(1-\beta)\delta}$$

  if $\mu_0 \geqslant \mu^\star$. In particular, $y_B = 0$ if $\mu = \mu^\star$.
- recommends N to all consumers if $\mu_0 < \mu^\star$.

Proof. First consider $-1 < a < 0$. The designer sets $y_B = 0$ in the second-best policy. By Proposition 9, there are only two candidate second-best policies: recommend Y until bad news, or recommend N to all consumers. Given $y_1 = 1$, $y_n = x_n$, by (15) and the law of motion of beliefs,

$$\begin{aligned}\sum_{n\geqslant 1}(2\mu_n - 1)y_n &= \sum_{n\geqslant 1}\delta^{n-1}\left[\mu_1 - (1-\mu_1)(1-\beta)^{n-1}\right]\\ &= \mu_0\left[\frac{1}{1-\delta} + \frac{1}{1-(1-\beta)\delta}\right] - \frac{1}{1-(1-\beta)\delta} \equiv \Gamma,\end{aligned}$$

so that ICC-Y is

$$-y_B + \Gamma \geqslant 0. \tag{19}$$

If $\Gamma \geqslant 0$, (19) is satisfied at $y_B = 0$. Also, since $\sum_{n\geqslant 1}[(1-a)\mu_n + a]y_n > \sum_{n\geqslant 1}(2\mu_n - 1)y_n = \Gamma$, recommending Y until bad news achieves strictly positive expected payoff (instead of 0 by always recommending N). Therefore, the second-best policy is to recommend Y until bad news arrives. If $\Gamma < 0$, (19) is violated even if $y_B = 0$. This indicates the only incentive feasible policy is to recommend N to all consumers, which is the designer's second-best policy.

Next, consider $a \geqslant 0$. y_B is no longer automatically 0 since the designer still gets a positive profit from a consumer purchasing a known bad product. If $\mu_0 < \mu^\star$, the argument is the same as the cases of $-1 < a < 0$. When $\mu^\star \leqslant \mu_0 < 1/2$, while the designer always wants to recommend Y given $a \geqslant 0$, ICC-Y binds, and y_B is at most $-\Gamma$. When $\mu_0 \geqslant 1/2$, the designer can recommend Y to all consumers. The posterior belief remains at $\mu_0 = 1/2$, so ICC-Y is satisfied. □

Figure 2 illustrates the designer's first and second-best policies for $-1 < a < 0$. Both policies take the form of recommending N forever when the prior is sufficiently low, and recommending

Y until bad news for higher priors. The first best and the second-best policies misalign only at intermediate priors, namely, $\mu_0 \in (\underline{\mu}, \mu^\star)$. That is, either the first best is IC, when $\mu_0 \notin (\underline{\mu}, \mu^\star)$, or there is no IC recommendation policy that induces any experimentation, when $\mu_0 \in (\underline{\mu}, \mu^\star)$. This offers a stark contrast: a product with a moderate initial belief of good quality can be recommended for a long period and gain popularity as long as there is no negative feedback, while no experimentation is possible with a slightly lower initial belief.

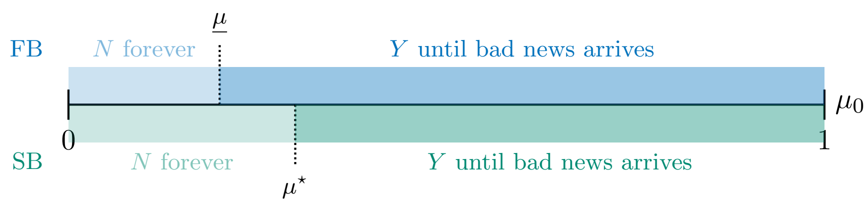

Figure 2 The first- and second-best policies for possible priors, negative a

When $a \geqslant 0$, the designer's optimal policies are illustrated in Figure 3. In the first-best policy, $\underline{\mu} = 0$, and Y is always recommended. However, recommending Y to all consumers is only incentive compatible if $\mu_0 \geqslant 1/2$. $\mu^\star$ again gives the boundary between recommending N forever and recommending Y to some consumers. For $\mu_0 \in (\mu^\star, 1/2)$, the designer always recommends Y as long as there is no bad news and occasionally recommends Y after the arrival of bad news so that the consumers are exactly indifferent between purchasing or not after observing a recommendation Y. At μ^*, the designer recommends Y until bad news arrives, then recommends N forever. When $\mu_0 < \mu^*$, the best he can do is to recommend N forever.

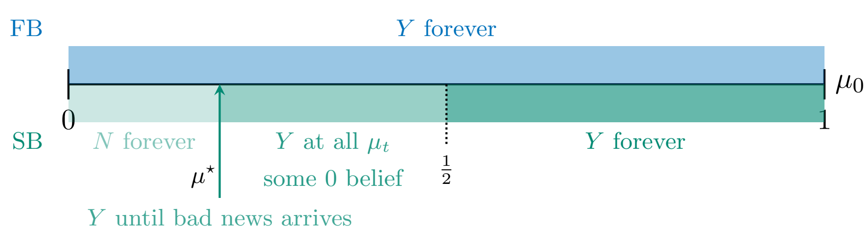

Figure 3 The first- and second-best policies under different priors, positive a

The comparative statics are similar to the perfect good news case: the designer is more willing to experiment when a, β and δ are higher, and the consumers are more willing to accept the recommendation when β and δ are higher. We summarize the designer's optimal policy for different a and μ_0 in Figure 4.

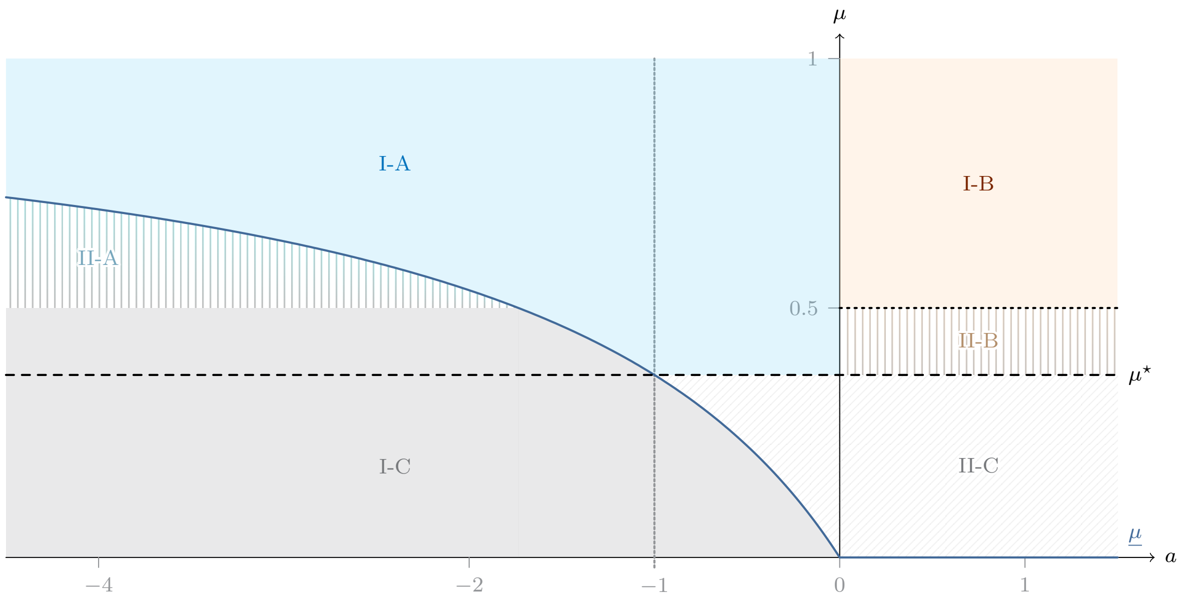

Figure 4 The second-best policy under perfect bad news

Note: Drawn at $\beta = 0.6, \delta = 0.55$. The designer can implement the first-best policies in the I regions (shaded with solid colors), while the first-best policies cannot be implemented in the II regions (shaded with line patterns). Color gray represents the policy that recommends N to all agents (Regions I-C, II-C), blue represents the policy that recommends Y until bad news arrives (Region I-A), and orange represents the policy that recommends Y to all consumers. In regions with vertical patterns (Regions II-A and II-B), the second-best policies involve mixing.

7 Extensions

7.1 Noisy Signal on Arrival Time

Until now we assumed that the only information a consumer receives is the recommendation generated by the system that the designer commits to. In a lot of settings though consumers might have some information about how recently the product was introduced to the market, so how early they are as adopters. Here we consider an extension, in the context of the simple example from subsection 2.1, in which consumers also receive a noisy signal about which period they arrive.

Suppose that consumers in the 2-period example in subsection 2.1, besides receiving a recommendation from the system whether to purchase the product, observe a signal k about their arrival time, where for some $p > 1/2$,

$$k(t) = \begin{cases} t & \text{with probability } p, \\ 3 - t & \text{with probability } 1 - p, \end{cases}$$

together with the designer's recommendation. We would like to check if the designer's first best is still implementable. Suppose the designer recommends Y to the first consumer. The designer recommends Y to the second consumer only if good news arrives. The beliefs of the consumers are

$$\mathbb{P}(t=1 \mid k=1, Y) = \frac{\frac{1}{1+\delta}p}{\frac{1}{1+\delta}p + \frac{\delta}{1+\delta}(1-p)\alpha\mu_0} = \frac{p}{p+\delta(1-p)\alpha\mu_0}.$$

$$\mathbb{P}(t=1 \mid k=2, Y) = \frac{\frac{1}{1+\delta}(1-p)}{\frac{1}{1+\delta}(1-p) + \frac{\delta}{1+\delta}p\alpha\mu_0} = \frac{1-p}{(1-p)+\delta p\alpha\mu_0}.$$

These allow us to write the IC constraints for consumers who receive $k=1$ and $k=2$.

$$\frac{p}{p+\delta(1-p)\alpha\mu_0}(2\mu_0-1) + \frac{\delta(1-p)\alpha\mu_0}{p+\delta(1-p)\alpha\mu_0} \times 1 \geqslant 0,$$
$$\frac{1-p}{(1-p)+\delta p\alpha\mu_0}(2\mu_0-1) + \frac{\delta p\alpha\mu_0}{(1-p)+\delta p\alpha\mu_0} \times 1 \geqslant 0,$$

which simplify to

$$\mu_0 \geqslant \frac{1}{2+\delta\alpha\left(\frac{1-p}{p}\right)} \quad \text{and} \quad \mu_0 \geqslant \frac{1}{2+\delta\alpha\left(\frac{p}{1-p}\right)},$$

respectively. Notice that $p > 1/2$, hence

$$\frac{1}{2+\delta\alpha\left(\frac{1-p}{p}\right)} \equiv \mu_p > \underline{\mu}_0 = \frac{1}{2+\delta\alpha} > \frac{1}{2+\delta\alpha\left(\frac{p}{1-p}\right)},$$

where $\underline{\mu}_0$, defined in (1), is the lowest belief that the designer recommends Y in the first period under the first-best policy. In other words, the binding IC is the IC of consumer who receives signal $k=1$.

Therefore, the designer's first best is no longer implementable if

$$\mu_0 \in \left[\underline{\mu}_0, \frac{1}{2+\delta\alpha\left(\frac{1-p}{p}\right)}\right) = \left[\underline{\mu}_0, \mu_p\right).$$

Inside this range, if the designer keeps the recommendation plan unchanged, only consumers who receive $k=2$ follow the recommendation Y. Such consumers believe that they are more likely to be the second consumer, who is recommended to purchase only when the product is revealed to be good. Consumers who receive $k=1$ instead believe that it is more likely that they arrived in the first period and are asked to experiment a product of unknown quality. This example shows that once consumers have additional information about their arrival time, the designer may not be able to implement the first-best policy as often even if their per-period payoff coincides with the consumers' payoffs. Notice that the bound $\mu_p < 1/2$ changes intuitively with p. As $p \to 1/2$, consumers receive imprecise information regarding the timing of their arrival, and $\mu_p \to \underline{\mu}_0$. At the extreme $p = 1/2$, consumers have no information about their arrival time, and the original first-best policy is recovered with $\mu_p = \underline{\mu}_0$. Conversely, as $p \to 1$, consumers become increasingly aware that they are asked to experiment, and $\mu_p \to 1/2$. In the limiting case $p = 1$, the first consumer knows with certainty that the designer possesses no additional private information; here, $\mu_p = 1/2$, reflecting the fact that no experimentation can be induced.

Another feature of this example is that the designer may be able to experiment even if the original first-best policy (when the consumers have no information about their arrival time) is not IC. More specifically, signal $k = 2$ is beneficial for the designer as it strengthens the confidence of the consumer that more likely she is not asked to experiment but to exploit the known information. If the first consumer receives signal $k = 2$, the designer can induce some experimentation at a belief even lower than $\underline{\mu}_0$. Generally speaking, the signal about the arrival time breaks the homogeneous consumer setup and creates types of consumers with different beliefs regarding arrival time. Each type has a different posterior regarding their arrival time, leading to a different set of IC constraints, which reflect different trade-offs between exploration and exploitation. The IC constraints for consumers who believe they arrive early based on their arrival time signal usually require a higher prior to accept recommendation Y as they believe that they are more likely to be asked to explore rather than exploit, while consumers who believe that they are more likely to arrive later can accept Y at a lower prior.

7.2 Some Consumers not Leaving Feedback

In this subsection we consider an extension of the model in which we relax the assumption that all consumers leave feedback about their experiences with the product. We focus on the case of a nonpaternalistic designer and a perfect good news signal structure.

Formally, for each consumer purchase, we assume that with probability $\kappa \in (0, 1)$ the consumer does not leave feedback for the system. For simplicity we assume that the signal that the consumer receives after purchasing the product and whether the consumer leaves a feedback are independent. Therefore if a consumer does not leave feedback, the belief regarding the product's quality remains the same.

The possibility of no feedback introduces exogenous delay to experimentation. That is, while every purchase leads to belief jumping to 1 or updating to a lower belief if feedback is always available, an experiment now can lead to no change in belief. Such delay slows down learning and reduces the value of experimentation. In turn, the designer stops experimenting at a higher belief. We show this by solving the designer's first-best problem.

First, by applying the same argument in the proof of Theorem 1, we can show that the designer's first-best policy is still a threshold policy. Let μ_κ be the threshold. For $\mu \geqslant \mu_\kappa$, the designer's Bellman equation is

$$W(\mu) = 2\mu - 1 + \delta \left\{ \kappa W(\mu) + (1-\kappa) \left[\frac{\mu\alpha}{1-\delta} + (1-\mu\alpha) W\left(\frac{\mu(1-\alpha)}{1-\mu\alpha} \right) \right] \right\},$$

which simplifies to

$$(1-\delta\kappa)W(\mu)=2\mu-1+\delta(1-\kappa)\left[\frac{\mu\alpha}{1-\delta}+(1-\mu\alpha)W\left(\frac{\mu(1-\alpha)}{1-\mu\alpha}\right)\right].$$

At μ_κ, $W(\mu_\kappa)$ and $W(\mu_\kappa(1-\alpha)/(1-\mu_\kappa\alpha))$ are both 0, so that

$$\mu_\kappa=\frac{1-\delta}{2(1-\delta)+\delta\alpha(1-\kappa)}.$$

Comparing this to $\underline{\mu}$ from Example 1, since $(1-\kappa)\in(0,1)$, $\mu_\kappa<\underline{\mu}$. That is, the designer stops experimentation at a strictly higher belief.

It is also straightforward to show that this first-best policy is IC given that $a=-1$. Intuitively, consumers understand that feedback is not always provided and adjust the belief update after observing a recommendation Y, which again aligns the payoff of consumers after observing a recommendation Y and the designer's expected payoff before Period 1, $W(\mu_0)$. We can show this formally using occupancy measures. Notice that the on path beliefs are the same as before, only the occupancy measure for each belief is different. In particular, instead of (11), we now have

$$\begin{aligned}
x_1&=1+\delta[x_1-(1-\kappa)y_1],\\
x_{n+1}&=\delta(1-\kappa)(1-\alpha\mu_n)y_n+\delta\left[x_{n+1}-(1-\kappa)y_{n+1}\right],\\
x_G&=\delta x_G+\delta\sum_{n\geqslant 1}(1-\kappa)\alpha\mu_n y_n.
\end{aligned}$$

Since $\mu_\kappa<1/2$, it is sufficient to check that ICC-Y is satisfied. From the law of motions above, both the designer's objective and the consumer's expected payoff after observing Y under the first-best policy (i.e., the LHS of ICC-Y) are

$$\sum_{n\geqslant 1}\left[\mu_n\left(2+\frac{\alpha(1-\kappa)\delta}{1-\delta}\right)-1\right]y_n.$$

Since the designer chooses to experiment only if the expected payoff of experimenting is weakly positive, this is weakly positive for $\mu_0>\mu_\kappa$. Therefore, ICC-Y is satisfied.

Just as in subsection 5.1, the coefficient of y_n can be considered as the value of experimentation at belief μ_n, by considering both the current period payoffs and all possible continuation payoffs. For each n, the coefficient here is smaller than the corresponding one in subsection 5.1 due to the additional $(1-\kappa)$ factor, which shows that the experimentation value is lower at each belief due to the possibility of no feedback and no learning.

8 Conclusion

Our results demonstrate that in an environment where agents do not know how many others previously experimented with a product, an information designer can have a lot of scope to encourage or discourage experimentation. While in environments in which agents know their orders of arrival, a designer, even when receiving private information from other sources, can only initiate experimentation slowly and gradually (Che and Hörner, 2018; Lyu, 2026), in the environments that we consider in some cases he can facilitate experimentation at the first-best level and intensity. In particular this is the case when the designer wants to maximize consumer surplus. If the designer's objective is sufficiently misaligned from this objective, then he cannot achieve his first best because of having to satisfy the constraints that his recommendations are credible enough. We leave it to future work to examine learning in similar information environments when there are multiple products of unknown quality from which to learn, and when there are different types of consumers with potentially different valuations for goods.

References

Hailemariam Ayalew, Jordan Chamberlin, and Carol Newman. Site-specific agronomic information and technology adoption: A field experiment from ethiopia. *Journal of Development Economics*, 156:102788, 2022.

Abhijit V Banerjee. A simple model of herd behavior. *The quarterly journal of economics*, 107(3): 797–817, 1992.

Dirk Bergemann and Stephen Morris. Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95, 2019.

Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, 100(5):992–1026, 1992.

Sushil Bikhchandani, David Hirshleifer, Omer Tamuz, and Ivo Welch. Information cascades and social learning. *Journal of Economic Literature*, 62(3):1040–1093, 2024.

Patrick Bolton and Christopher Harris. Strategic experimentation. *Econometrica*, 67(2):349–374, 1999.

Yeon-Koo Che and Johannes Hörner. Recommender systems as mechanisms for social learning. *The Quarterly Journal of Economics*, 133(2):871–925, 2018.

Alex Gershkov and Balázs Szentes. Optimal voting schemes with costly information acquisition. *Journal of Economic Theory*, 144(1):36–68, 2009.

Rema Hanna, Sendhil Mullainathan, and Joshua Schwartzstein. Learning through noticing: Theory and evidence from a field experiment. *The Quarterly Journal of Economics*, 129(3):1311–1353, 2014.

Nicole Immorlica, Jieming Mao, Aleksandrs Slivkins, and Zhiwei Steven Wu. Incentivizing exploration with selective data disclosure. *arXiv preprint arXiv:1811.06026*, 2018.

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101 (6):2590–2615, 2011.

Godfrey Keller and Sven Rady. Strategic experimentation with poisson bandits. *Theoretical Economics*, 5(2):275–311, 2010.

Godfrey Keller and Sven Rady. Breakdowns. *Theoretical Economics*, 10(1):175–202, 2015.

Godfrey Keller, Sven Rady, and Martin Cripps. Strategic experimentation with exponential bandits. *Econometrica*, 73(1):39–68, 2005.

Sergey Kovbasyuk and Giancarlo Spagnolo. Memory and markets. *Review of Economic Studies*, 91(3):1775–1806, 2024.

Ilan Kremer, Yishay Mansour, and Motty Perry. Implementing the "wisdom of the crowd". *Journal of Political Economy*, 122(5):988–1012, 2014.

Chen Lyu. Information design for social learning on a recommendation platform. *Journal of Economic Theory*, page 106150, 2026.

Chen Lyu and Lu Liao. Information design for social learning with patient agents. *PHBS Working Paper 20260301*, 2026.

Erik Madsen and Eran Shmaya. Collective upkeep. *Available at SSRN 4887505*, 2025.

Yishay Mansour, Aleksandrs Slivkins, Vasilis Syrgkanis, and Zhiwei Steven Wu. Bayesian exploration: Incentivizing exploration in bayesian games. *Operations Research*, 70(2):1105–1127, 2022.

Kaisa Miettinen. *Nonlinear multiobjective optimization*, volume 12. Springer Science & Business Media, 1999.

Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming.* John Wiley & Sons, 1994.

Luis Rayo and Ilya Segal. Optimal information disclosure. *Journal of political Economy*, 118(5): 949–987, 2010.

Max Simchowitz and Aleksandrs Slivkins. Exploration and incentives in reinforcement learning. *Operations Research*, 72(3):983–998, 2024.

Aleksandrs Slivkins. Exploration and persuasion. In Federico Echenique, Nicole Immorlica, and Vijay V. Vazirani, editors, *Online and Matching-Based Market Design*, pages 655–675. Cambridge Univ. Press Cambridge, UK, 2023.

Nancy L Stokey and Robert E Lucas. *Recursive methods in economic dynamics.* Harvard University Press, 1989.

A An Example of Multiplicity of Implementation

In this example, we show that two different policies can share the same occupancy measure, same value for the designer, and same incentive compatibility. Let $\mu_0 = \mu_1$ and α be given. Suppose $\delta = 4/5$ for concreteness. Consider the following two policies. Y is always recommended at belief 1.

- Policy 1: Recommend Y at μ_1 at Period 1. If no news arrives, recommend Y at μ_2 with probability 1/2 at Period 2. Recommend N from Period 3 onwards unless the belief is 1.
- Policy 2: Recommend Y at μ_1 at Period 1. If no news arrives, recommend Y at μ_2 with probability 2/5 at Period 2 and with probability 5/24 at Period 3. Recommend N at belief μ_3 at Period 3. Recommend N from Period 4 onwards unless the belief is 1.

The two policies share the same occupancy measure. In particular, $y_1 = 1$, $y_2 = \delta(1-\mu_1\alpha)/2$, and $y_k = 0$ for $k \geqslant 3$. The same y's suggest that x_G must be the same between two policies.

We first consider the designer's value, assuming that the consumers obey the recommendations. Under both policies, the values take the following form

$$(1-a)\mu_1 + a + \delta\mu_1\alpha\frac{1}{1-\delta} + \delta(1-\mu_1)\alpha \times [\text{continuation payoffs}].$$

More precisely, the continuation payoffs in the bracket are the continuation payoffs if the belief at the beginning of Period 2 is μ_2. Under Policy 1, this continuation payoff is

$$\frac{1}{2}\left[(1-a)\mu_2 + a + \mu_2\alpha\delta\frac{1}{1-\delta}\right].$$

And under Policy 2, this continuation payoff is

$$\frac{2}{5}\left[(1-a)\mu_2 + a + \mu_2\alpha\delta\frac{1}{1-\delta}\right] + \frac{3}{5} \times \delta \times \frac{4}{25}\left[(1-a)\mu_2 + a + \mu_2\alpha\delta\frac{1}{1-\delta}\right].$$

Substituting $\delta = 4/5$ into this expression, it is easy to see that the two continuation values are the same.

We next consider the incentive compatibility. Consider the expected belief after observing Y. Under Policy 1, a consumer observes Y at beliefs μ_1, μ_2, and 1, and the conditional belief after observing Y is the average of the three beliefs, each weighted by the probabilities of reaching that belief at a particular period and the probabilities of recommendation at that belief. But this means the calculation of the posterior belief rather similar to calculating the values above. For example, under Policy 1, the posterior belief after observing Y is

$$\frac{1}{P}\left(\mu_1 + \delta\mu_1\alpha\frac{1}{1-\delta} \times 1 + \delta(1-\mu_1)\alpha \times \frac{1}{2}\left[\mu_2 + \mu_2\alpha\delta\frac{1}{1-\delta} \times 1\right]\right),$$

where

$$P = 1 + \mu_1\alpha\frac{\delta}{1-\delta} + \frac{\delta}{2}(1-\mu_1\alpha)\left(1 + \mu_2\alpha\frac{\delta}{1-\delta}\right).$$

It is easy to verify that, given $\delta = 4/5$, the posterior belief after observing Y under Policy 2 coincides with this expression as well. By Bayesian consistency, the posterior belief after observing N must remain the same as well. Therefore, the two policies share the same incentive compatibility.